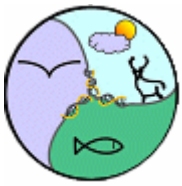
Climatic Change as an Engine for Speciation in Orthoptera Species:

Flightless Species on African Mountains as a Model System

Kjetil Lysne Voje

Master of Science thesis
2007

Department of Biology
University of Oslo



Centre for Ecological and
Evolutionary Synthesis

FORORD / PREFACE

Denne oppgaven ville på ingen måte blitt til uten hjelp og støtte fra en rekke personer. Først og fremst fortjener mine tre veiledere Nils Christian Stenseth, Øystein Flagstad og Glenn-Peter Sætre en stor takk. Ikke alle professorer våger å satse på en bachelor-student når et helt ukjent modellsystem på andre siden av ekvator skal utforskes, men Nils Christian gjør nettopp det. Det er jeg veldig glad og takknemlig for. Ved siden av all faglig og ikke-faglig støtte fortjener også Nils Christian en takk for å ha dannet et unikt spennende, dyktig og inkluderende forskningssenter, som det som masterstudent er veldig inspirerende å få være en del av. Øystein og Glenn-Peter har begge vært fantastiske støttespillere gjennom hele prosessen. Deres gode ideer, konstruktive og gjennomtenkte tilbakemeldinger på mine tekstutkast og ukuelige positive holdning til mitt arbeid har vært (og er fortsatt) høyt verdsatt. Det samme kan sies om mitt ikke-faglige samvær med dem begge, det være seg en øl på puben eller fullstendig utslitende Holmenkollstafetter. Takk også til Claudia og Andreas Hemp, som tok meg imot med åpne armer og viste meg sitt Afrika. Uten Claudia ville det neppe ha blitt spesielt mange gresshopper å analysere i Oslo. Takk også til alle på CEES-laben for et godt arbeidsmiljø og Lise Heier for konstruktiv hjelp og nyttig brainstorming rundt en av de statistiske analysene. Takk også til Hans Kristian Voje, Ragnhild Heimstad og Torbjørn Bruvik for korrekturlesing av oppgaven.

Av andre som har bidratt, både faglig og ikke fullt så faglig, må alle godtfolk tilknyttet lesesalen (også kjent som Centre for Scientific and Cultural Synthesis) fremheves. Ingen nevnt, ingen glemt, men dere skal vite at dere alle har bidratt enormt til min høye trivselsfaktor disse to årene. Det samme gjelder også i aller høyeste grad mine venner utenfor biologimiljøet. En stor takk også til mine kjære foreldre. Min glede og fascinasjon for naturen har jeg primært fra dere. Sist, men ikke minst, en spesiell takk til Ragnhild, for tålmodighet og engasjement, og for å ha gjort denne tiden helt fantastisk.

Kjetil Lysne Voje
Blindern, 31. mai 2007

CONTENTS

ABSTRACT

Climatic fluctuations characterized the climate in East Africa during the Plio-Pleistocene. These variable conditions established opportunities for ecological fragmentation with subsequent genetic isolation of species dependent on both savannah and forest habitats. This study addresses the hypothesis that historical climatic fluctuations in the region have had a major evolutionary impact on the fauna in East Africa. To do this, mitochondrial DNA sequences from a group of bush crickets consisting of both forest and savannah inhabiting taxa were analyzed in relation to Plio-Pleistocene range fragmentations indicated by palaeoclimatic studies. Parts of the COI and 12S mitochondrial genes were sequenced to infer phylogeographic patterns. Coalescent modeling and mismatch distributions were used to distinguish between alternative biogeographic scenarios. The same methods, together with a traditional molecular clock approach, were also used to date the divergence times between species. The results indicate that two radiations, one between 6 and 3.5 million years ago and the other about 0.8 million years ago, gave rise to most of the species analyzed. The earliest radiation overlaps in time with the global spread of C4 grasslands and speciation through adaptation to this new type of plants may therefore explain the contemporary sudden burst of savannah inhabiting lineages. Climate seems to have been the driving force behind the more recent of the two radiations. An intensified drying and cooling of the climate right after the last warm and wet maximum in the region about 1 million years ago resulted in the retraction of forest to higher altitudes. Forest dependent species were consequentially trapped in forest refuges and simultaneous vicariant speciation events in these lineages followed. Accordingly, this study shows that the East African continental archipelago is a suitable model system for phylogenetic research and for illuminating how climatic fluctuations may influence speciation and other evolutionary processes. Further, my results suggest that savannah inhabiting species have experienced intense selective constraints on the mtDNA COI gene compared to forest dependent species. One possible explanation could be severe competition on the savannah for the right microclimate and food resources. Finally, paraphyly at both the genus and the species level was detected and some taxonomic revisions are suggested.

INTRODUCTION

The term phylogeography was first used in a paper by Avise and colleagues (1987) and the subject is often seen as a subdicipline of the much older research tradition of biogeography. Typical for phylogeography is the emphasis on historical aspects of the contemporary spatial distribution of gene lineages (Avise 1998). It is an integrative research discipline in the way that analysis and interpretation of the spatial lineage distribution very often depend upon input from different areas of research, like molecular genetics, phylogenetics, population genetics, historical geography and geology. Most phylogeographic research has been conducted at the intraspecific level, often comparing the spatial and genealogical relationships between different geographical populations. However, the techniques are also applicable for studying certain systems of closely related species (e.g. Allegrucci *et al.* 2005; Trewick 2000; Trewick and Morgan-Richards 2005)

Island archipelagos are attractive model systems for phylogeographic analyses, in large part because geographic isolation of the islands restricts gene flow between populations on individual islands (Emerson and Hewitt 2005). Fauna and floral systems on the Hawaii, the Canary Islands and the Galapagos Islands have been studied extensively and researchers have used these archipelagos to address a range of evolutionary questions, both above and below the species level (e.g. Emerson and Oromi 2005; Gillespie 2004; Pestano *et al.* 2003; Tonnis *et al.* 2005). Ocean, together with distance, separate the islands in these archipelagos, creating barriers which reduce the exchange of genes between the islands. A quite analogous system can be found in East Africa where former and more recent geological activity have created forested volcano and mountain "islands" separated by an "ocean" of savannah over variable distances (Figure 1). Because of the current climate, forests are more or less absent in the lowland and are restricted to higher altitudes on the mountains (Figure 2). As a consequence, forests and their inhabitants have a very patchy distribution in East Africa. In this respect, these mountain systems are a continental version of the ocean archipelagoes and should

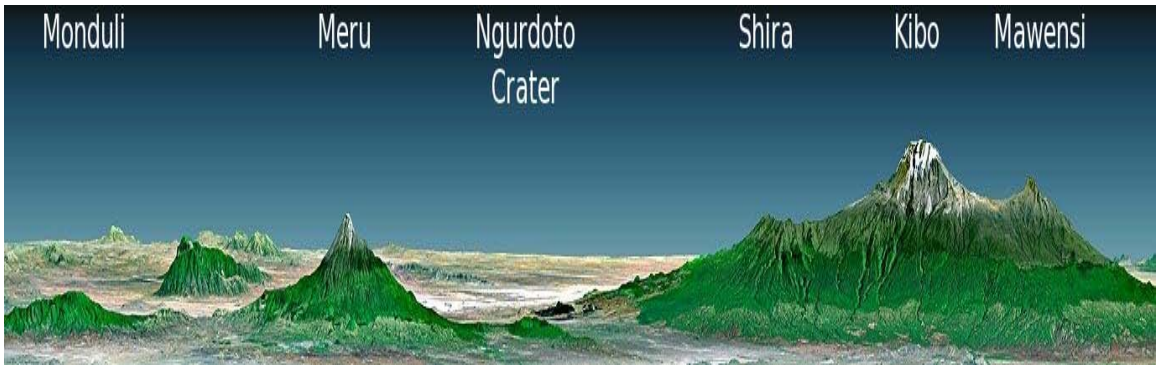


Figure 1 Some of the volcanoes and mountains in East Africa. Forests are restricted to certain altitudes and are not found on the plains separating the volcanoes and mountains. Image Credit: NASA/JPL/NIMA

therefore be potentially well suited for doing phylogeographic research. Yet, very few phylogeograpic studies have been conducted on this continental archipelago system.

One apparent difference between the oceanic and continental archipelagoes is the greater possibility of forest patches to expand and retract under different climatic regimes compared to oceanic islands, making the inhabitants of forest islands to a greater extent only temporarily isolated from each other. The temporal distribution of forests through history may therefore have played a crucial role in shaping the phylogeography of species dependent on forest habitats. Fortunately, a lot of palaeoclimatic research has been conducted in Africa, especially in the eastern part of the continent. Apart from the intrinsic scientific interest of palaeoclimate, one reason for this massive research on historical climate in this particular region is the fact that climatic change is believed to have had an important influence on evolution and community change in mammals in general and in hominids in particular (see e.g. Dennell 2003; Finlayson 2005). Because *Homo* and its ancestors have undergone significant evolutionary change the last five million years, the geological epochs of Plio- and Pleistocene have been particularly well studied with respect to climatic history (Fernandez and Vrba 2006). These studies confirm a continuing drying trend in the region, but also recurrent climatic fluctuations of wet and warm and dry and cold climatic regimes respectively (deMenocal 1995; 2004; Fernandez and Vrba 2006). In the wet and warm periods, it is assumed that the forest cover expanded down from the mountains to lower altitudes. Accordingly, originally isolated forest patches may have become connected during these periods, allowing allopatric forest inhabiting populations to meet and potentially mix and spread. During

dry and cold periods, the forest probably retracted to higher altitudes and was replaced by savannah landscape between the mountains, very much like the situation is at present (Figure 3).

Important ecological events other than the cyclic variation in forest cover have also taken place in East Africa over the last million years. During the late Miocene and early Pliocene a massive increase in C4 grasslands occurred in large parts of the globe (Cerling *et al.* 1993; Morgan *et al.* 1994; Latorre *et al.* 1997; Barry *et al.* 2002). The transition from a C3 dominated vegetation to a mainly C4 based vegetation gave rise to several new habitats and must have posed a major challenge for all species dependent on the ousted C3 plants. Further, the East African rift zone has given rise to several new volcanoes over the last million years, including Mt. Kilimanjaro and Mt. Meru (Lovett and Wasser 1993). Novel habitats at different altitudes on young volcanoes are therefore

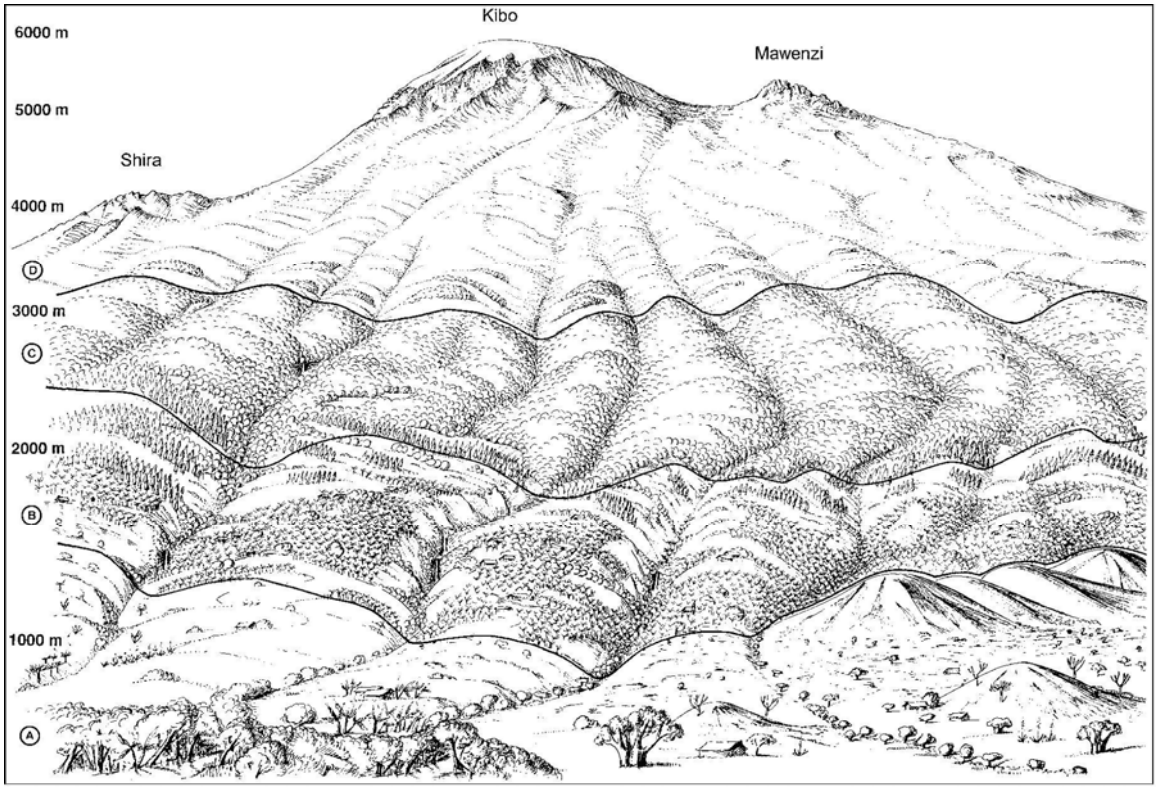


Figure 2 A schematic outline of the distribution of different types of landscape at the southern slope of Mt. Kilimanjaro. A: colline zone, B: submontane zone, C: montane zone, D: alpine zone. Forest is restricted to higher altitudes (B and C) and a typical savannah landscape is found at lower altitudes (A). Originally published in Hemp (2002a).

just another example of the ecological innovations and alterations East Africa has experienced over the last million years.

My primary goal with this study is to investigate how the recent and major ecological events mentioned above might have influenced speciation, radiation and dispersal (i.e. the evolutionary history) of insects belonging to East Africa's biota. To do this I use a group of bush-crickets (Orthoptera: Tettigoniidae), consisting of both forest dependent species and savannah inhabiting species as a model system. As a consequence of the patchy distribution of forest in the region are many of the forest dependent species restricted to one particular mountain or volcano in the region. Another typical feature is these species' reduced wings (Ragge 1960) and, as a consequence, diminished migration potential (Hemp 2005). Together, these characteristics make this group of insects a good candidate to help me reach my second main goal; to investigate the suitability of the

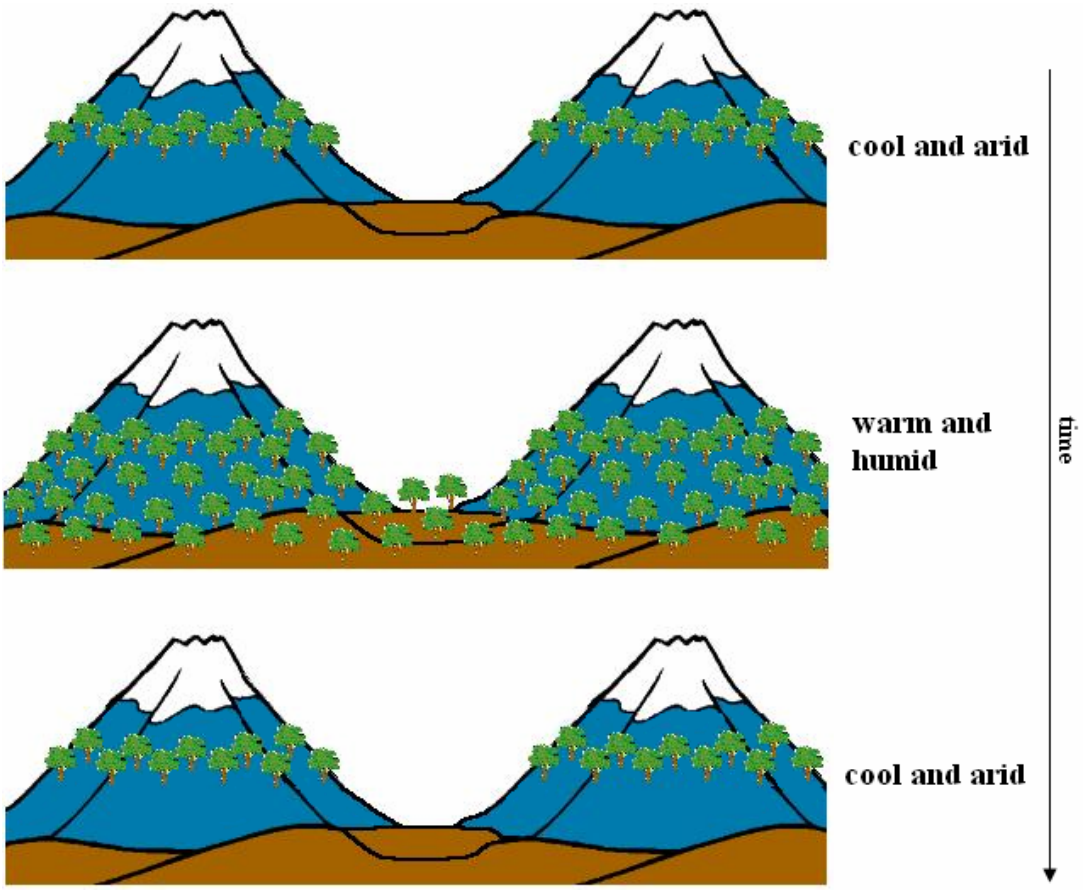


Figure 3 Schematic representation of range shifts of forest in response to climatic fluctuations during the Plio-Pleistocene. Forests were captured at high altitudes during cool and arid climatic conditions while the forest expanded downward when the climate became more wet and warm. Temporary forest bridges connecting originally isolated forest patches may have originated during these warmer periods.

continental "island archipelago" of East Africa as a model system for general assessment of the role of climatic impact on speciation and other evolutionary processes.

The bush crickets analyzed represents four different genera which all belong to the subfamily Phaneropterinae. The genus *Monticolaria* consists solely of forest dependent species while the genus *Horatosphaga* comprises a mix of species inhabiting forest and savannah habitats, respectively. Species belonging to the genera *Lamecosoma* and *Peronura* are all savannah inhabitants. While the forest species in this study are often endemic to a single mountain or volcano the savannah species have a much more widespread distribution. Information about all species analyzed is shown in Table 1. In the text I refer to forest dependent species as those only found at forest edges or within forests. Four species previously not described are included in the study. These have been given preliminary names, indicated by brackets in Table 1, which are consistently used throughout the text.

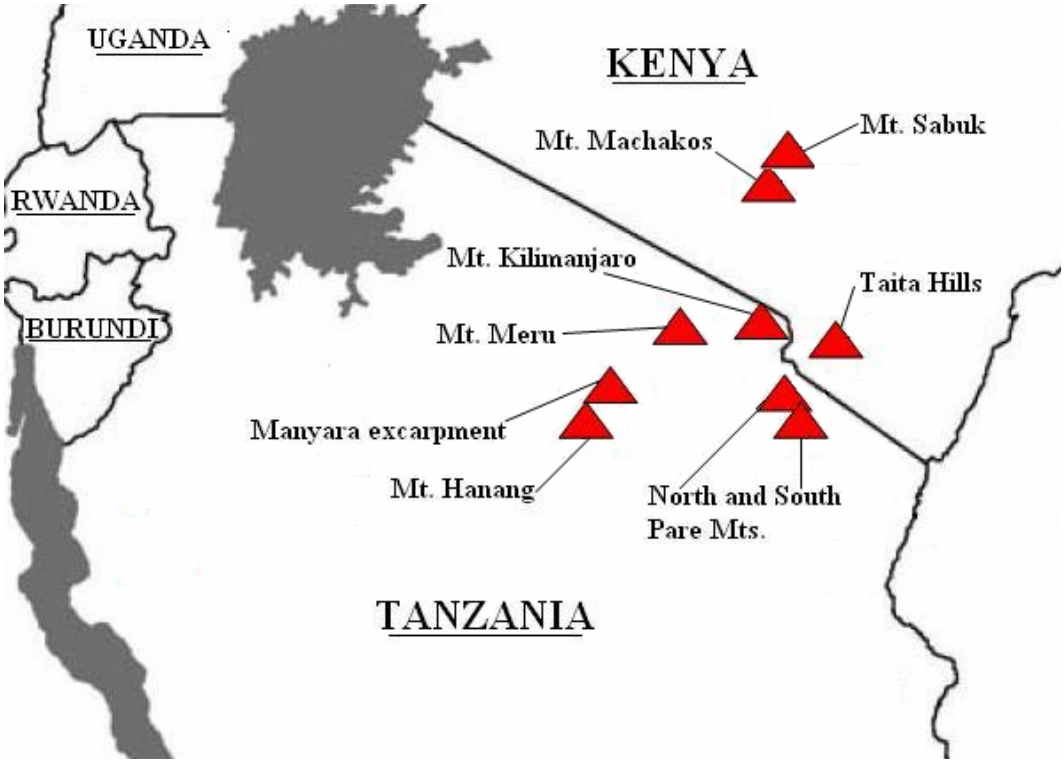


Figure 4 Map showing approximate location of mountains, volcanoes and other geological formations inhabited by species included in the study.

The mitochondrial (mtDNA) genome has several favorable qualities for studies of evolutionary history (Avise *et al.* 1987), such as maternal inheritance, lack of recombination and absence of complicated genetic structures like repetitive DNA, transposable elements, pseudogenes and introns. Moreover, the mtDNA genome contains a number of genes with different evolutionary rates. I therefore sequenced two different mitochondrial segments, one corresponding to a fragment of the small subunit of the ribosomal RNA (12S) and the other to the subunit I of cytochrome oxidase (COI), to reconstruct phylogenetic relationships and to assess the evolutionary history of this group of bush crickets.

Several hypotheses are addressed in this study. First, as forest seems to predate savannah in East Africa, I hypothesize that savannah adopted lineages evolved from forest living species. The spread of C4 grasslands at the expense of C3 vegetation about 7 to 4 million years ago probably formed opportunities for adaptive speciation to the new radiating plant variants. My next hypothesis is therefore that savannah inhabiting lineages started to originate around the boundary between the Miocene and the Pliocene. I also hypothesize that savannah lineages have experienced a history with less rigid selective constraints compared to what forest dependent lineages have experienced, as a relaxed selection regime might have been necessary to make the habitat transition possible. Further, climatically induced ecological fragmentation of forest with subsequent genetic isolation of forest species on mountains should have created ample opportunities for vicariant speciation in forest dependent lineages. However, speciation through adaptation to novel habitats on geologically young volcanoes might also explain much of the observed biodiversity in forest living bush crickets in the region. In an attempt to explore the driving force behind speciation in the forest dependent lineages I therefore address both the vicariant and the adaptive speciation hypotheses. Finally, as the phylogeny of this group is unresolved, I also look for monophyly of the different genera and discuss to what extent taxonomic revisions are needed.

Table 1 Summary of current range, region, collection locations and habitat type of the species included in the study. Arrows indicate that species were collected in the region given in the table. The abbreviation n.sp mean "new species".

Species	Current range	Region	collection location (number of specimens)	Habitat type	Source Refs.
Horatosphaga heteromorpha Karsch, 1888	widespread	East Africa	North Pare(1), Kilimanjaro(1), Taita Hills(2), Uluguru(1), Meru(1), Turtle(1), W-Usambara(1)	colline to montane bush- and grassland	Hemp 2005
Horatosphaga (hanangensis) n.sp.	narrow/endemic	Mt. Hanang, Tanzania	← (5)	clearings of indigenous submontane forests	Hemp pers.comm.
Horatosphaga nou Hemp, 2006	narrow/endemic	Manyara Escarpment, Tanzania	← (3)	herbaceous vegetation along montane forest edges and clearings	Hemp 2006
Horatosphaga parensis Hemp, 2002b	narrow/endemic	North and South Pare Mts.	Nort Pare only(2)	herbaceous vegetation along montane forest edges and clearings	Hemp 2002b
Horatosphaga (tenera) n.sp.	narrow	Ngong Hills, Kenya	← (3)	savannah grasslands	Hemp pers.comm.
Horatosphaga regularis Bolivar, 1922	widespread?	Kenya	Ngong Hills(4) & Naivasha Escarpment(2)	submontane savannah grasslands	Ragge 1960
Horatosphaga montivaga Sjöstedt, 1909	narrow/endemic	Mt. Meru & Mt. Kilimanjaro, Tanzania	← (7)	clearings of indigenous submontane forests	Hemp and Hemp 2003
Horatosphaga meruensis Sjöstedt, 1909	narrow/endemic	Mt. Meru & Mt. Kilimanjaro, Tanzania	← (1)	colline and submontane grasslands	Hemp 2006
Horatosphaga sabuk Hemp, 2006	narrow	Highlands on Mt. Machakos & Mt. Sabuk, Kenya	← (4)	herbaceous vegetation along montane forest edges and clearings	Hemp 2006
Lamecosoma inermis Ragge, 1961	widespread?	Tanzania, Kenya	Kilimanjaro(3)	colline and submontane savannah	Hemp pers.comm.
Lamecosoma (new) n.sp	narrow?	Mt. Hanang, Tanzania	← (5)	montane grassland	Hemp pers.comm.
Peronura clavigera Karsch, 1888	widespread	Kenya, Tanzania	North Pare(4), Kilimanjaro(6), Taita Hills(3)	herbaceous vegetation mostly along forest edges, montane grassland	Hemp pers.comm.
Peronura uguenoensis Hemp, 2002b	narrow/endemic	North Pare Mts., Tanzania	← 10	grassland intermingled with herbs, forest edges	Hemp 2002b
Monticolaria kilimandjarica Sjöstedt, 1909	narrow/endemic	Mt. Kilimanjaro, Tanzania	← (4)	disturbed submontane forests and disturbed and undisturbed montane forests	Hemp and Hemp 2003
Monticolaria meruensis Sjöstedt, 1909	narrow/endemic	Mt. Meru, Tanzania	← (3)	disturbed submontane forests and disturbed and undisturbed montane forests	Hemp pers.comm.
Monticolaria (hanangensis) n.sp.	narrow/endemic	Mt. Hanang and Nou forest, Tanzania	← (3)	disturbed and undisturbed montane forests and forest edge	Hemp pers.comm.

MATERIALS AND METHODS

Sampling

The analyzed specimens were collected between 1999 and 2006, either by hand or by the use of catching nets. Most samples were stored in 75-90% ethanol but some were stored dry prior to extraction of DNA. Sample sizes are for the most part small, reflecting the rarity of many of the species. However, whenever possible I analyzed multiple individuals to provide an indication of the genetic diversity within species. All species determinations were done by Claudia Hemp. In all instances, appropriate authorization to collect and use specimens was obtained from the Commission for Science and Technology of Tanzania.

DNA Extraction, PCR Amplification and Sequencing

I removed one hind leg from each specimen for DNA extraction and cut it into small pieces. DNA was extracted using a standard phenol-chloroform protocol (Sambrook ***et al.*** 1989). Segments of the mitochondrial genes cytochrome oxidase I (COI) and 12S were amplified using polymerase chain reaction (PCR). The primers SR-J-14233 (5′ AAG AGC GAC GGG CGA TGT GT 3′) and SR-N-14588 (5′ AAA CTA GGA TTA GAT ACC CTA TTA T 3′) (Simon ***et al.*** 1994) were used in amplifying segments of the third domain of the 12S gene while LCO1490 (5′ GGT CAA CAA ATC ATA AAG ATA TTG G 3′) and HCO2198 (5′ TAA ACT TCA GGG TGA CCA AAA AAT CA 3′) (Folmer ***et al.*** 1994) were used to amplify a fragment of the COI gene.

PCR was performed in 20µL volumes containing 0.60 µL of each primer (10µM), 200µM dNTPs, $MgCl_2$ between 2.0-3.0µM and 0.6 units of HotStarTaq DNA polymerase (QIAGEN). I used 2µL of DNA extract for amplification. A touch-down program was used and the annealing temperature started at 55°C and was reduced by 1°C in the next 4 cycles. The rest of the 30-35 cycles comprised denaturation at 94°C for 1 minute, annealing temperature at 50°C for 30 seconds and extensions at 72°C for 1.5 minutes. The number of total cycles in the PCR protocol varied due to variable amplification success among samples. I screened all PCR products on a 2% agarose gel stained with

ethidium bromide. PCR products were then cleaned with ExoSAP-IT (USB Corporation, USA) following the manufacturer's recommendations, except for the amount of ExoSAP-IT used per μL PCR product (ExoSAP-IT was diluted 10 times) and the incubation time for degradation of primers and nucleotides (37°C for 30 minutes instead of 15 minutes). Sequencing was performed on an ABI 3730 high-throughput capillary electrophoresis instrument. The same primers used in PCR were also used in sequencing. I aligned COI sequences by eye. The 12S sequences were aligned by ClustalW (MEGA v.3.1 (Kumar *et al.* 2004)) and inspected and revised manually.

Phylogenetic Analyses

I examined phylogeographic patterns by inferring phylogenetic relationships from mitochondrial sequence data and by comparing these relationships with the species' present distribution. In total, I sequenced 567 bp from 80 individuals and 405 bp from 34 individuals from the COI and 12S genes respectively. Although genes in the non-recombining mitochondrial genome of animals are not independently evolving loci, they often exhibit considerable heterogeneity of evolutionary rate (Simon *et al.* 1994) and may therefore sometimes yield incongruent phylogenetic estimates (e.g. Wetzer 2002). I therefore ran the Partition-Homogenity Test (also known as the Incongruence Length Difference Test) (Farris *et al.* 1994; 1995) as implemented in PAUP* v.4.0b10 (Swofford 2001) to check if the data from the two mtDNA fragments could be combined to build a robust phylogeny. Starting trees were obtained via stepwise addition with random additions of taxa using 10 replicates. Tree bisection-reconnection (TBR) was chosen as the branch swapping algorithm using 10,000 generations.

No indication of incongruent phylogenetic estimates were found so I combined the two data sets and made a larger one (COI/12S) comprising 972 bp from 34 ingroup individuals. The 12S data set did not contain enough genetic variation to reveal reliable phylogenetic signals. Accordingly, I did not compute phylogenetic trees based solely on the 12S sequences. However, I analyzed the COI data set also separately because of the larger sample size compared to the combined COI/12S data set. All phylogenetic analyses utilized a homologue COI and 12S sequence of *Drosophila melanogaster* from Genbank as an outgroup (GeneID: 192469 and 261003 in GenBank). I did not find any Orthoptera

species outside the subfamily Phaneropterinae in GenBank which had been sequenced for both the COI and 12S sequences used in this study. However, the genetically distinct genus *Monticolaria* functions as an outgroup to the rest of the ingroup in addition to *D. melanogaster*.

To define which parameters to be used in the different phylogenetic analyses I used ModelTest v3.7 (Posada and Crandall 1998). This software package compares different models of nucleotide evolution in a hierarchal hypothesis testing framework using both log likelihood values and the Akaike information criterion (AIC) (Akaike 1973). A General Time Reversible model with invariable sites and a gamma distribution (GTR + I + Γ) best fitted both the COI and the COI/12S data set.

Bayesian inference (BI) analysis (Huelsenbeck and Ronquist 2001) on both data sets was completed in MrBayes v.3.1.2 (Huelsenbeck and Ronquist 2001). I set the priors to match the model chosen by Modeltest but did not fix any of the parameters. Two independent analyses were run simultaneously, each starting from different random trees. Each search was run with four Markov chains for 2,000,000 and 1,000,000 generations for the COI and COI/12S data set respectively. Trees were sampled every 1,000 generation. I discarded the first 500,000 generations as burn-in for the COI dataset, creating a dataset consisting of 1,501 trees for each independent run. For the combined dataset, I discarded the first 250,000 generations as burn-in, creating a dataset of 751 trees for each independent run. Plots of the generation versus the log probability of the data given the parameter values were analyzed for both datasets to ensure the two independent analyses had reached stationarity before the end of the burn-in. I also examined tree topologies from individual Markov chains to check for consistency of results. The proportion of searches in which any given node is found during the Markov chain Monte Carlo process, constitutes the Bayesian posterior probability for that node.

To examine the robustness of the topology identified in the BI analyses I also computed phylogenetic trees using neighbor-joining (NJ) (Saitou and Nei 1987), maximum parsimony (MP) (Farris 1970) and maximum likelihood (ML) (Felsenstein 1981) on both data sets. I implemented the GTR + I + Γ model before starting the NJ and ML tree building processes. NJ analyses were preformed in PAUP* while I used the freely available University of Oslo Bioportal (http//:www.bioportal.uio.no/) to conduct

the ML analyses. Starting-trees prior to the ML analysis were obtained from random stepwise addition with 10 replicates followed by TBR. MP analyses were conducted in MEGA. Searches were performed using close-neighbor-interchange with search level 1 and the initial tree was generated by a random addition of sequences with 10 replications. I examined support for nodes in all NJ, ML and MP analyses using bootstrap with 1,000 replications, except for the ML analysis of the COI data set where I used 100 replications to save time.

Estimation of Divergence Rate

The timing of divergences between species and populations provides an important test of phylogeographic hypotheses. In this respect, the time since they split apart must be consistent with the putative historical events that are hypothesized to have driven their divergence. In order to estimate when lineages diverged, however, it is necessary to have an accurate estimate of the evolutionary rate in which mutations accumulate in the DNA sequences used.

Using COI consensus sequences made for each species in BioEdit v.5.0.9 (Hall 1999), I applied Tajima's relative rate test (Tajima 1993) to check for substitution rate variation between species. As implemented in MEGA, I conducted pairwise comparisons using both the 1D (all substitutions combined) and 2D (transitions and transversions considered separately) methods. Of the 240 pairwise species comparisons conducted, only 9 indicated that the molecular clock hypothesis had to be rejected. *H. tenera* was included in all 9. Genetic distances between species were then calculated using PAUP* and the same COI consensus sequences. Both corrected (using the GTR + I + Γ model) and uncorrected distances (p-distances) were computed. The reliability of such genetic distances often decreases with increased sequence divergence due to multiple substitutions. I therefore plotted transitions (TIs) and transversions (TVs) versus corrected genetic distance to search for mutation saturation using COI consensus sequences (Figure 5). The plot of TIs vs. genetic distances shows a small tendency for TI saturation for genetic distances greater than 0.2. No such saturation is detected in the plot of TVs vs. genetic distances. Accordingly, nucleotide substitutions seem to accumulate linearly up to a genetic distance of approximately 0.2.

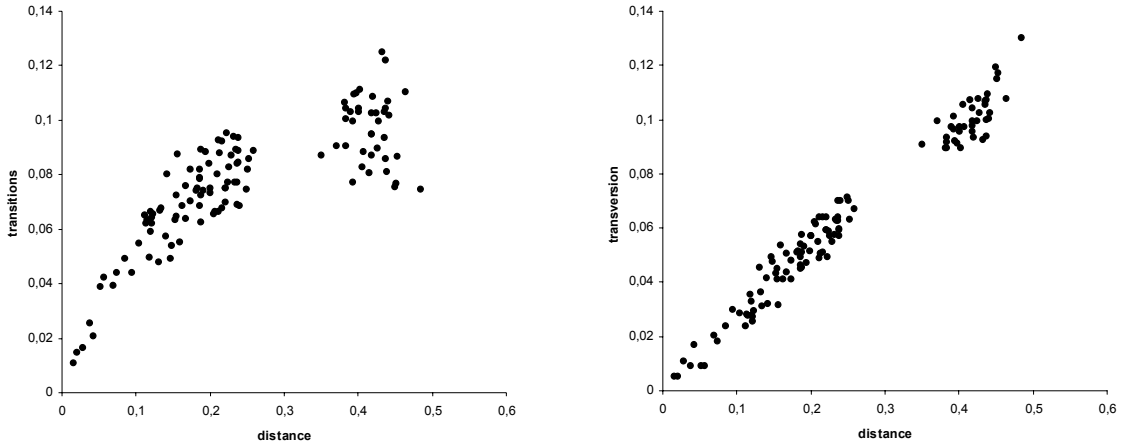


Figure 5 Transitions (left) and transversions (right) versus corrected genetic distance (using a GTR + I + Γ model) for the COI data set. Each species is represented with one consensus sequence.

Molecular divergence rates applicable to the mitochondrial genome can be obtained from previous studies where the divergence time between clades has been calibrated with, for example, geological events (Brower 1994; Juan *et al.* 1995; 1996; Fleischer *et al.* 1998). I find this approach useful, but not all together satisfactory since the divergence rates estimated in the mentioned studies vary from 1.2 to 2.3 % per lineage, per million years. These discrepancies between previously estimated divergence rates emphasize the need of estimating the evolutionary rate in each study separately if some sort of calibration is possible, if only to investigate if the divergence rate seems to be about the same as found in previous studies. Fortunately, ages of certain volcanoes inhabited by species included in this study have been roughly dated by geologists. Mt. Kilimanjaro and Mt. Meru are both assumed to be the result of the same geological events in the Rift Valley about 1.000.000 years ago (Lovett and Wasser 1993). These two volcanoes are inhabited by *M. kilimandjarica* and *M. meruensis* respectively, while *H. montivaga* exists on both volcanoes. As these species cannot have existed on these particular volcanoes for a longer period of time than the actual geological age of the volcanoes, this might make it possible to infer the timing of speciation events.

Distinguishing Between Alternative Biogeographic Hypotheses and Dating of Speciation Events

Distinguishing between alternative biogeographic hypotheses is often difficult. Here, I used two methods to differentiate between different historical biogeographic scenarios. In order to distinguish historical isolation from ongoing migration I applied the coalescent approach developed by Nielsen and Wakeley (2001) using the MDIV software package produced by the same researchers. The second method, a mismatch distribution approach (Slatkin and Hudson 1991; Rogers and Harpending 1992), was used to test for recent historical population expansions in groups identified in the phylogenetic trees which showed signals consistent with such a scenario (e.g. groups with polytomies and where short branch lengths separated species or clades). The mismatch distribution approach is normally used on samples from different populations to test for, and estimate the timing of, intraspecific population expansion. However, the approach can also work on groups that have experienced recent speciation events since species radiation within a short period of time to a large extent is genetically equivalent to a regular intraspecific population expansion (Sturmbauer *et al.* 2005).

The stochastic nature of the coalescent process often makes it difficult to make reliable estimates of the divergence time between two lineages. Further, every approach for estimating divergence times has certain shortcomings. Accordingly, one should try to use more than one method in order to investigate the reliability of the estimate found. In this study I therefore use both the coalescent and the mismatch distribution approach discussed above to estimate the divergence times between some of the recent lineages in my data set. The two approaches differ a lot from each other in how they calculate parameters. Congruence in time estimates found by the two methods may therefore indicate that the estimates can be considered reliable. Ancestral and old lineage splitting events do often not conform to the assumptions made by the models applied in these two approaches. The timing of old speciation events were therefore estimated by the use of a traditional molecular clock approach.

In the coalescent analyses, I conducted four pairwise comparisons on five different species belonging to two different monophyletic clades. Each pairwise analysis included species separated by low genetic distances, indicating relatively recent speciation events. A detailed description of Nielsen and Wakeley's (2001) method and

how I performed the coalescent analyses in MDIV is presented in Appendix 1. The mismatch distributions were calculated as implemented in ARLEQUIN v.2.000 (Schneider *et al.* 2000). Haplogroups showing branch length patterns that could indicate recent population expansions were analyzed. A detailed description of the approach is available in Appendix 2.

Non-Synonymous versus Synonymous Substitutions

Different selective regimes will often leave distinct and detectable genetic signatures in lineages which have experienced these regimes. One such signature is the ratio of non-synonymous versus synonymous substitutions in a protein coding DNA segment. For example, strong purifying selection tends to allow only synonymous substitutions to happen in a coding DNA segment. Accordingly, species which have experienced strong selective constraints might show a lower ratio of non-synonymous versus synonymous substitutions for a particular coding DNA segment compared to species which have experienced less rigid stabilizing selection. To test for different historical selection regimes, I computed the number of non-synonymous substitutions per non-synonymous site (*Ka*) and the number of synonymous substitutions per synonymous site (*Ks*) for any pair of sequences within three monophyletic groups in the COI phylogeny. I conducted the computations in the program DNAsp (Rozas *et al.* 2003). Using estimates of Pi (π) (the average of the *Ka* and *Ks* values of all pairwise comparisons within taxa) given by the same program, I computed the average ratios of non-synonymous versus synonymous substitutions (dn/ds) within each group. I used a bootstrap approach conducted in R v.2.4.0 (R Development Core Team 2006) to test if the dn/ds ratios were significantly different from each other. Two-tailed 99% confidence intervals for the difference between the dn/ds ratios, using 10,000 replications, were computed.

RESULTS

Phylogenetic Analyses

The COI and COI/12S trees obtained from the Bayesian inference (BI) analyses are shown in Figure 6 and 7, respectively. Both trees reflect the same phylogenetic relationships. The inter-species relationship within the monophyletic genus *Monticolaria* are completely resolved, *M. hanangensis* and *M. meruensis* being the closest relatives. *P. uguenoensis*, *P. clavigera* and *H. regularis* form another monophyletic group and signals in my data indicate that *H. regularis* and *P. clavigera* are sister groups with *P. uguenoensis* forming the sister group to this two-species clade. Strong support for monophyly was also found for the species *H. montivaga, H. nou* and *H. hanangensis*. Further, the species *L. new, L. inermis, H. heteromorpha, H. meruensis* and *H. tenera* do also form a monophyletic group. Notably though, within this group the intraspecific lineages of *L. inermis* and *H. heteromorpha* do not form monophyletic clades in the COI tree. However, both these species are monophyletic in the combined BI tree, which highlight the importance of sequencing more than a few samples from each species.

All the monophyletic groups found in the BI analyses mentioned above are confirmed by the neighbor joining (NJ), maximum parsimony (MP) and maximum likelihood (ML) analyses (Appendix 3). The rest of the topology found in the BI analyses is also largely congruent with the results from these three additional tree building approaches.

In general, the more recent clades in the evolutionary trees have high clade support. This is not always the situation for some of the deeper parts of the phylogenies. The order of old lineage splitting events may therefore be difficult to judge in some cases, as is evident from the polytomy seen in the combined BI tree.

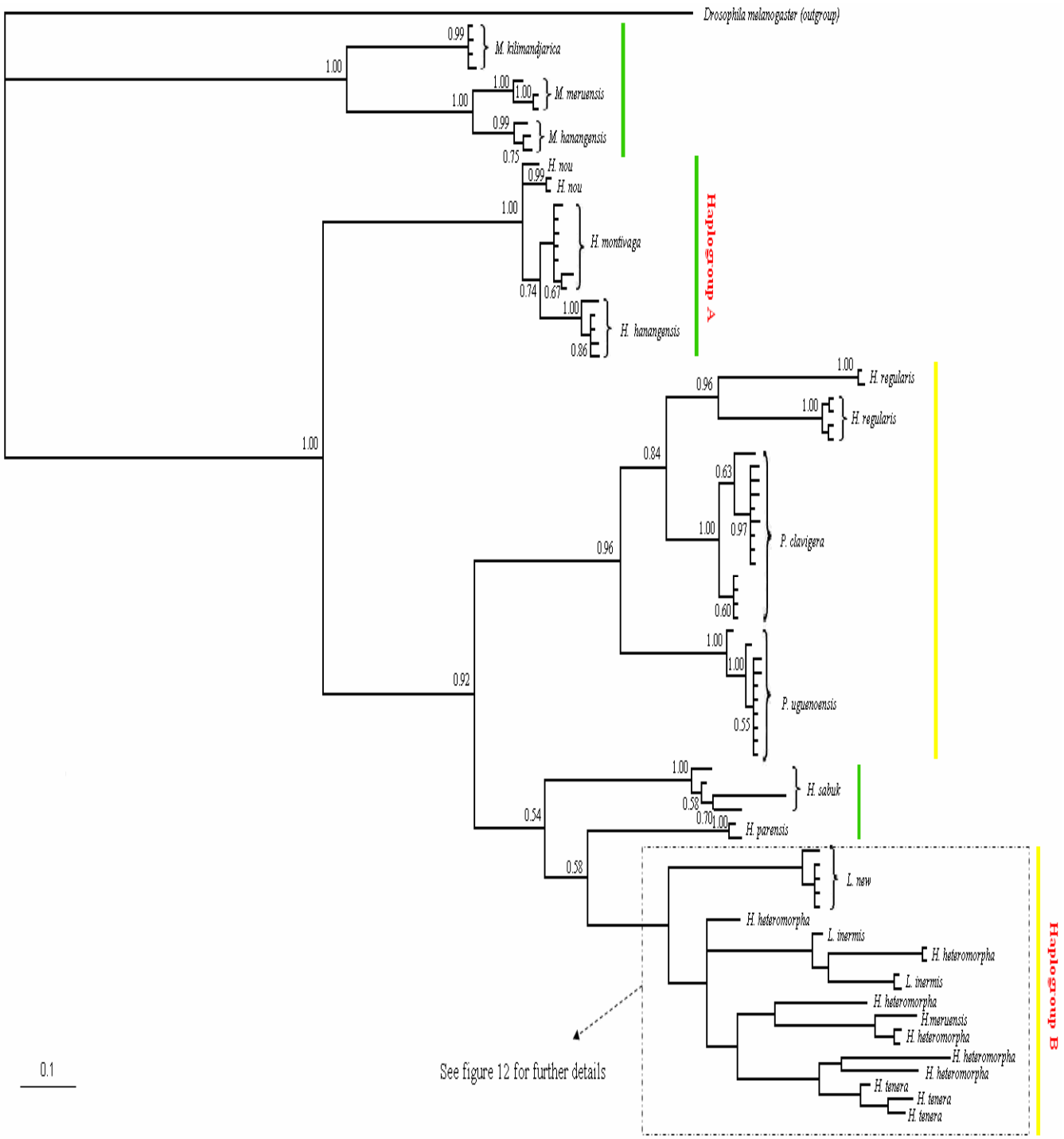


Figure 6 Bayesian inference of the phylogeny of the species included in the study, based on the COI-gene. Priors were set to match a GTR + I + Γ model. *Drosophila melanogaster* was used as an outgroup. Branch lengths are unconstrained with an exponential distributed prior and support values are indicated for nodes that were supported in 50 % or more of the sampled trees. The arithmetic mean of the marginal likelihood of the sampled trees is – 4262.46. Colors indicate habitat type: green = forest dependent species, yellow = savannah inhabiting species. Haplogroups analyzed in the mismatch distribution approach are also indicated.

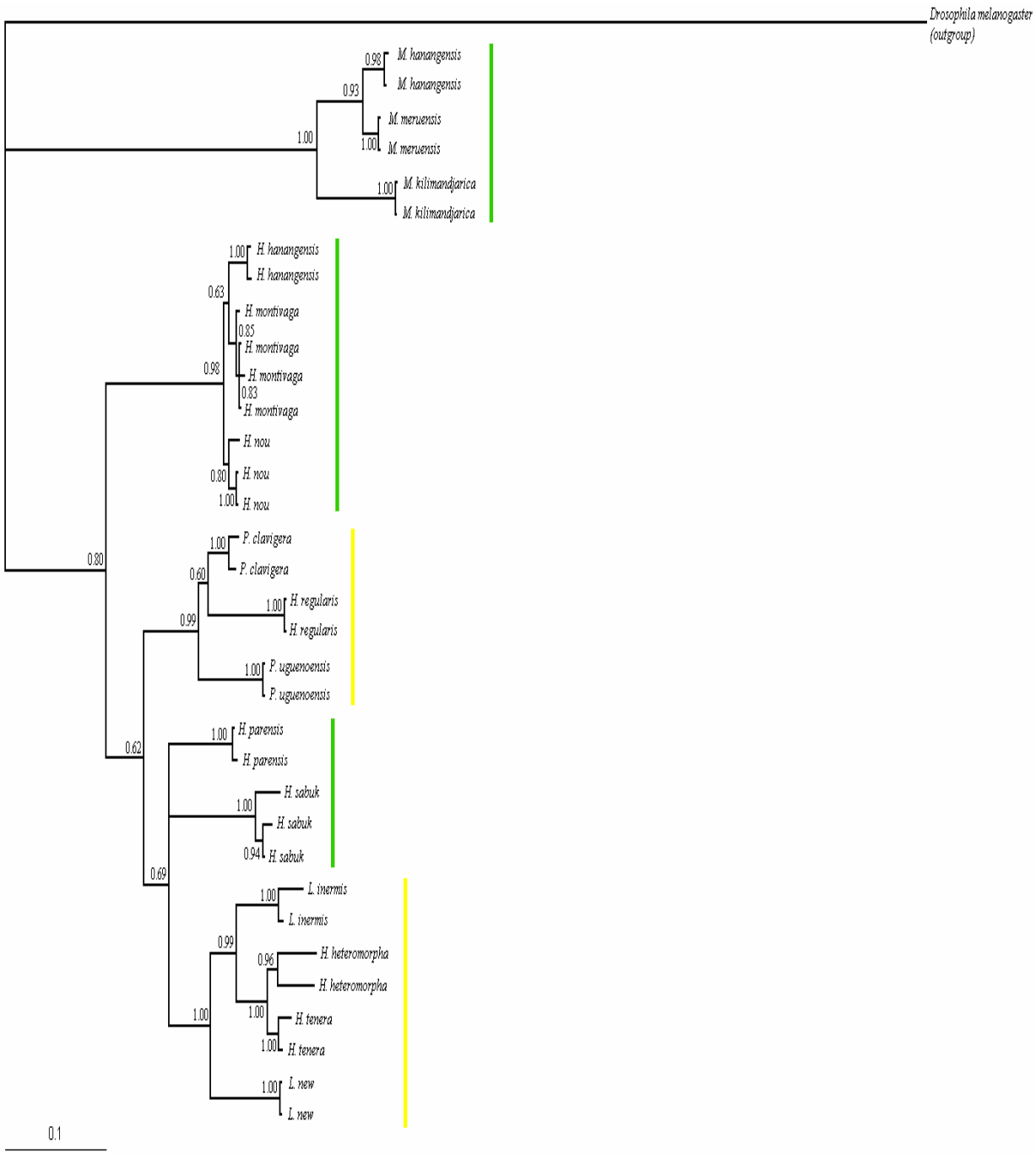


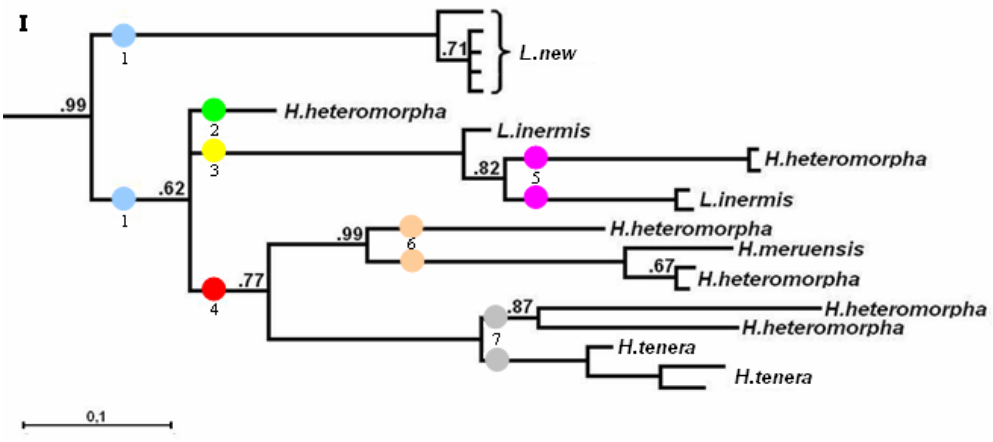
Figure 7 Bayesian inference of the phylogeny of the species included in the study, based on the combined COI/12S dataset. Priors were set to match the model chosen by Modeltest, i.e. allowing gamma-distributed rate variations across sites and a certain proportion of invariable sites. *Drosophila melanogaster* was used as an outgroup. Branch lengths are unconstrained with an exponential distributed prior and support values are indicated for nodes that were supported in 50 % or more of the sampled trees. The arithmetic mean of the marginal likelihood of the sampled trees is – 5207.90. Colors indicate habitat type: green = forest dependent species, yellow = savannah inhabiting species.

Divergence Rate

To obtain a plausible divergence rate I considered previously reported substitution rates for the COI gene. These reported values were compared with genetic distances belonging to species restricted to the dated volcanoes Mt. Kilimanjaro and Mt. Meru. All genetic distances computed are shown in Appendix 4.

H. montivaga, living on both Mt. Kilimanjaro and Mt. Meru, has a genetic distance of 1.6 and 2.0% to *H. nou* and *H. hanangensis*, respectively. *M. kilimandjarica* and *M. meruensis*, both named after the volcanoes they inhabit, have larger genetic distances to their closest species-relative. While *M. meruensis* is 3.8% genetically different from *M. hanangensis*, *M. kilimandjarica* has a genetic distance of about 12% to both *M. meruensis* and *M. hanangensis.* An evolutionary rate of 2% divergence between lineages per million years has often been found when calibrating the COI divergence against geological data (Brower 1994; Juan *et al.* 1995; 1996). Thus, to obtain a 12% genetic divergence in the COI gene in only one million years seems very improbable, suggesting *M. kilimandjarica* lived elsewhere before Mt. Kilimanjaro was formed. The other three genetic distances (1.6, 2.0 and 3.8%) are closer to previously reported divergence rates, although 3.8% per lineage per million years also appears unusually long. A 2% divergence rate has often been used by researchers using the COI gene in phylogeographic studies on insects (Trewick and Morgan-Richards 2005; Spooner and Ritchie 2006), and seems reasonable to use also in my case.

The low support of some of the old lineages observed in the phylogenetic trees is not desirable from a taxonomic point of view. More data are probably needed to better understand the lineage splitting events in these parts of the phylogeny. However, low clade support can also suggest a radiation of new species originating more or less simultaneously. To investigate this hypothesis I computed genetic distances and their equivalents in million years (using a molecular clock approach) for some old lineage splitting events found in the phylogenetic analyses (Figure 8). Because of low clade support I also estimated divergence times between some of the main lineages in the group containing the paraphyletic savannah inhabiting lineages (see I in Figure 8). As the figure shows, most lineages analyzed seem to have



II	2	3	4
2		**0.082** (min: 0.060, max: 0.094)	**0.079** (min: 0.062, max: 0.089)
3	**0.061** (min: 0.053, max: 0.064)		**0.118** (min: 0.081, max: 0.153)
4	**0.061** (min: 0.047, max: 0.096)	**0.075** (min: 0.062, max: 0.109)	
1	GTR + I + Γ = **0.119** (min: 0.085, max:0.141), p - distance = **0.081** (min: 0.055, max: 0.070)		
5	GTR + I + Γ = **0.070** (min: 0.069, max:0.071), p - distance = **0.047** (min: 0.047, max: 0.047)		
6	GTR + I + Γ = **0.080** (min: 0.079, max:0.083), p - distance = **0.056** (min: 0.055, max: 0.058)		
7	GTR + I + Γ = **0.066** (min: 0.055, max:0.078), p - distance = **0.044** (min: 0.036, max: 0.051)		

III	2	3	4
2		**4.10 million years**	**3.95 million years**
3	**3.05 million years**		**5.90 million years**
4	**3.05 million years**	**3.75 million years**	
1	GTR + I + Γ = **5.95 million years**, p - distance = **4.05 million years**		
5	GTR + I + Γ = **3.50 million years**, p - distance = **2.35 million years**		
6	GTR + I + Γ = **4.00 million years**, p - distance = **2.80 million years**		
7	GTR + I + Γ = **3.30 million years**, p - distance = **2.20 million years**		

IV			
	H. sabuk & H. parensis	GTR + I + Γ = **0.119**	p - distance = **0.072**
	P. clavigera & P. uguenoensis	GTR + I + Γ = **0.086**	p - distance = **0.059**
	P. clavigera & H. regularis	GTR + I + Γ = **0.074**	p - distance = **0.045**
	H. regularis & P. uguenoensis	GTR + I + Γ = **0.134**	p - distance = **0.076**
	M. kilimandjarica & M. meruensis	GTR + I + Γ = **0.121**	p - distance = **0.074**

V			
	H. sabuk & H. parensis	GTR + I + Γ = **5.95** million years	p - distance = **3.60** million years
	P. clavigera & P. uguenoensis	GTR + I + Γ = **4.30** million years	p - distance = **2.95** million years
	P. clavigera & H. regularis	GTR + I + Γ = **3.70** million years	p - distance = **2.25** million years
	H. regularis & P. uguenoensis	GTR + I + Γ = **6.70** million years	p - distance = **3.80** million years
	M. kilimandjarica & M. meruensis	GTR + I + Γ = **6.05** million years	p - distance = **3.70** million years

Figure 8 A section of the Bayesian inference tree (I) based on the COI data showing the different lineage splitting events (indicated by both numbers and colors) for which genetic distances (II) and divergence times (III) have been estimated using a molecular clock (2% divergence between lineages per million years). The distances between lineages 2 (green), 3 (yellow) and 4 (red) above the diagonal are corrected using a GTR + I + Γ model. Distances below the diagonal are uncorrected (p-distances). Some genetic distances (IV) and divergence times (V) of old lineage splitting events are also shown. The genetic distances in II and the divergence times in III have been calculated using the mean distances between the lineages. The values in IV and V were estimated from each species' consensus sequence.

originated between 3.5 and 6 million years ago. The great majority of these species are savannah inhabiting lineages.

Coalescent Analyses

To date the more recent speciation events I used coalescent modeling. The same coalescent approach is also used in an attempt to distinguish historical isolation from ongoing migration. Five different species belonging to two different monophyletic clades were analyzed.

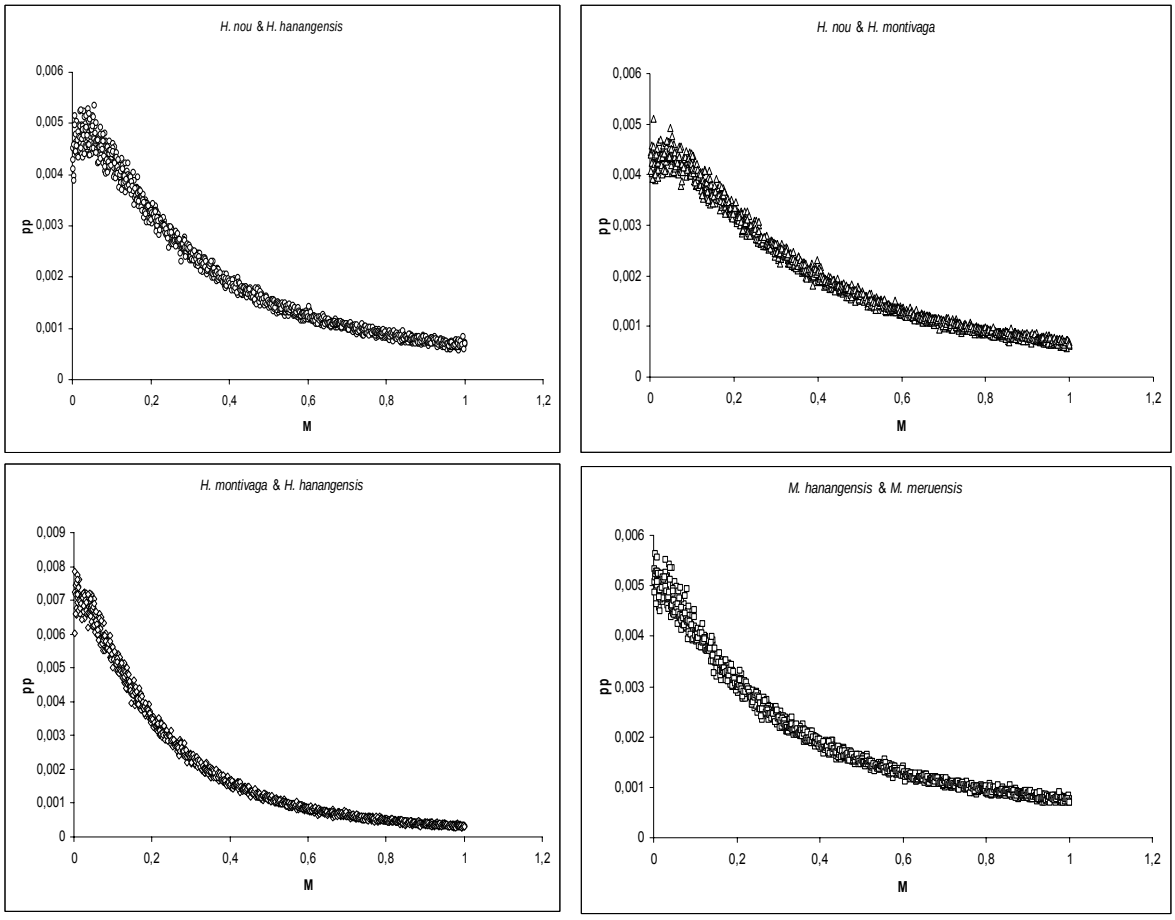


Figure 9 Three independent replicates of the posterior probabilities of migration (M) for each of the four pairwise species comparisons. Five species belonging to two different monophyletic clades were analyzed Upper left: *H. nou & H. hanangensis*, lower left: *H. montivaga & H. hanangensis*, upper right: *H. nou & H. montivaga,* lower right: *M. hanangensis & M. meruensis.*

For all four pairwise comparisons, estimates of migration rates suggested zero or extremely low levels of recent gene flow between the species compared (Figure 9).

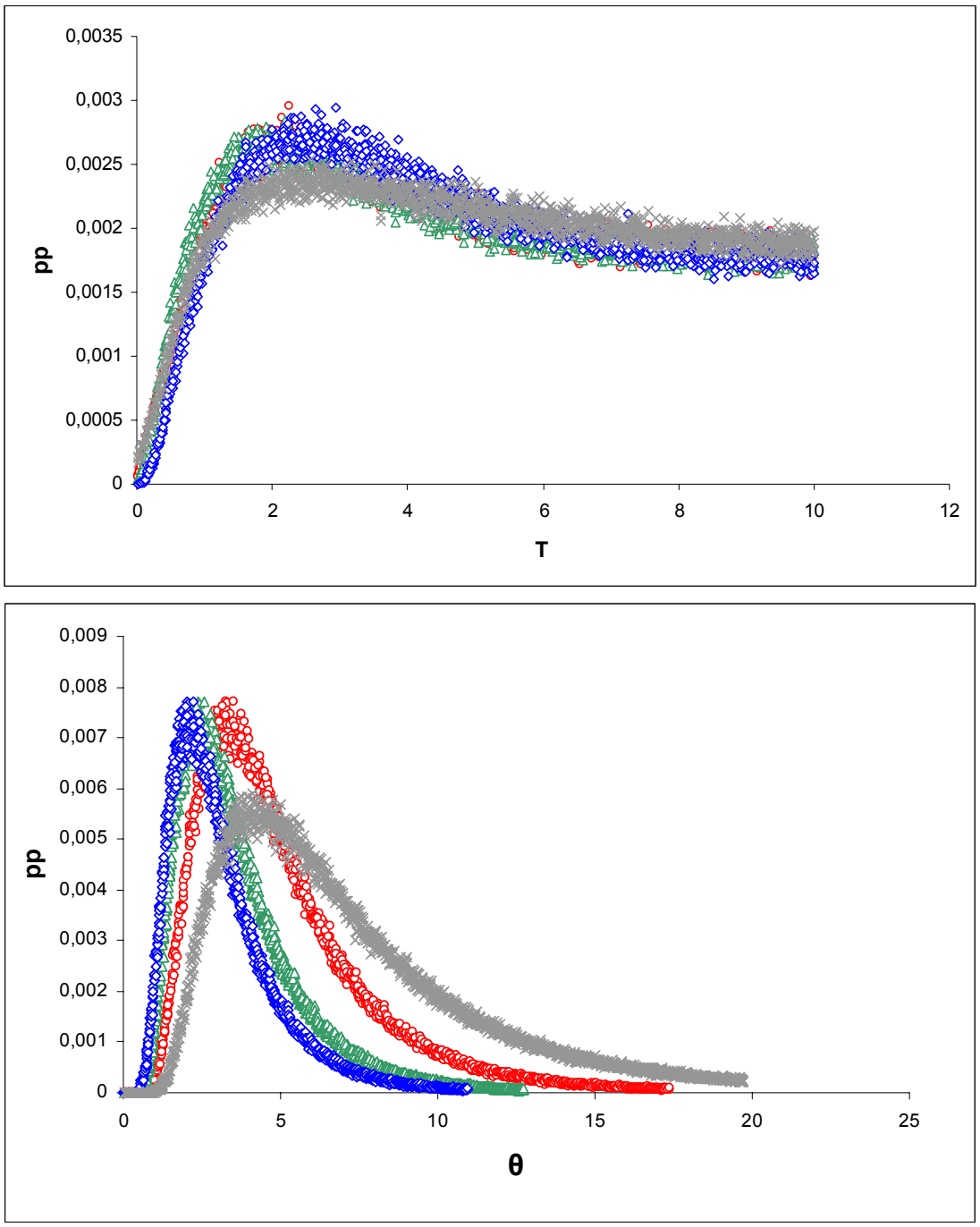


Figure 10 Three independent replicates of the posterior probabilities (pp) for divergence time (T) (upper) and for theta (θ) (lower) for each of the four pairwise comparisons between species; ○ (red) = *H. nou* & *H. hanangensis*, ◊ (blue) = *H. montivaga* & *H. hanangensis*, Δ (green) = *H. nou* & *H. montivaga*, × (grey) = *M. hanangensis* & *M. meruensis*

Accordingly, I did not calculate confidence intervals for M (M, θ and T are defined in Appendix 1). The posterior distributions of both divergence time and θ show a very large degree of overlap in all pairwise comparisons (Figure 10). The point estimates of T are very similar and ranges from 2.20 between *H. nou* and *H. montivaga* and 2.94 between *H. hanangensis* and *H. montivaga*. The point estimates of θ range from 1.95 between *H. hanangensis* and *H. montivaga* (CI = 0.70 – 5.36) and 4.10 (CI = 1.54 – 15.14) between *M. meruensis* and *M. hanangensis*. All posterior distribution curves for T decreases very slowly towards zero. For this reason I did not calculate confidence intervals for T either. Parameter estimates of T and θ were used to estimate divergence times between the species (T_{div}) (Table 2). The three *Horatosphaga* species seem to have originated during a time period lasting no more than 200,000 to 250,000 years.

By definition within coalescent theory, the time to common ancestry (tMRCA) must always be longer than the time since divergence (T_{div}) between two populations. This is confirmed by the values for tMRCA computed (Table 2). I found *H. montivaga* and *H. hanangensis* to have a lower tMRCA value than *H. nou* and *H. montivaga*, although the estimated T_{div} values suggest that *H. montivaga* diverged later from *H .nou* than from *H. hanangensis.* Such discrepancies might indicate that these three species originated more or less at the same time.

Table 2 Estimates of time since divergence (*T*) and time to most recent common ancestor (*tMRCA*) among four pairwise comparisons given in years before present (ybp). *Lower* and *upper* time since divergence represent the extreme values which lie within 2 AIC unites (ΔAIC) on either side of the point estimate of *T*.

	Tdiv (ybp)	lower (ybp)	upper (ybp)	tMRCA (ybp)
H. nou & H. hanangensis	698,683	235,200	2,026,874	938,630
H. nou & H. montivaga	464,132	162,940	1,268,957	713,085
H. montivaga & H. hanangensis	504,288	181,317	1,388,209	645,197
M. hanangensis & M. meruensis	925,408	347,028	3,416,892	1,379,911

Mismatch Distributions

To further date lineage splitting events and to check for historical population expansions, I used a mismatch distribution approach.

Two peaks can be seen in the observed mismatch distribution of haplogroup A (*H. nou*, *H. montivaga* and *H. hanangensis*) (bars in Figure 11, I), a pattern not compatible

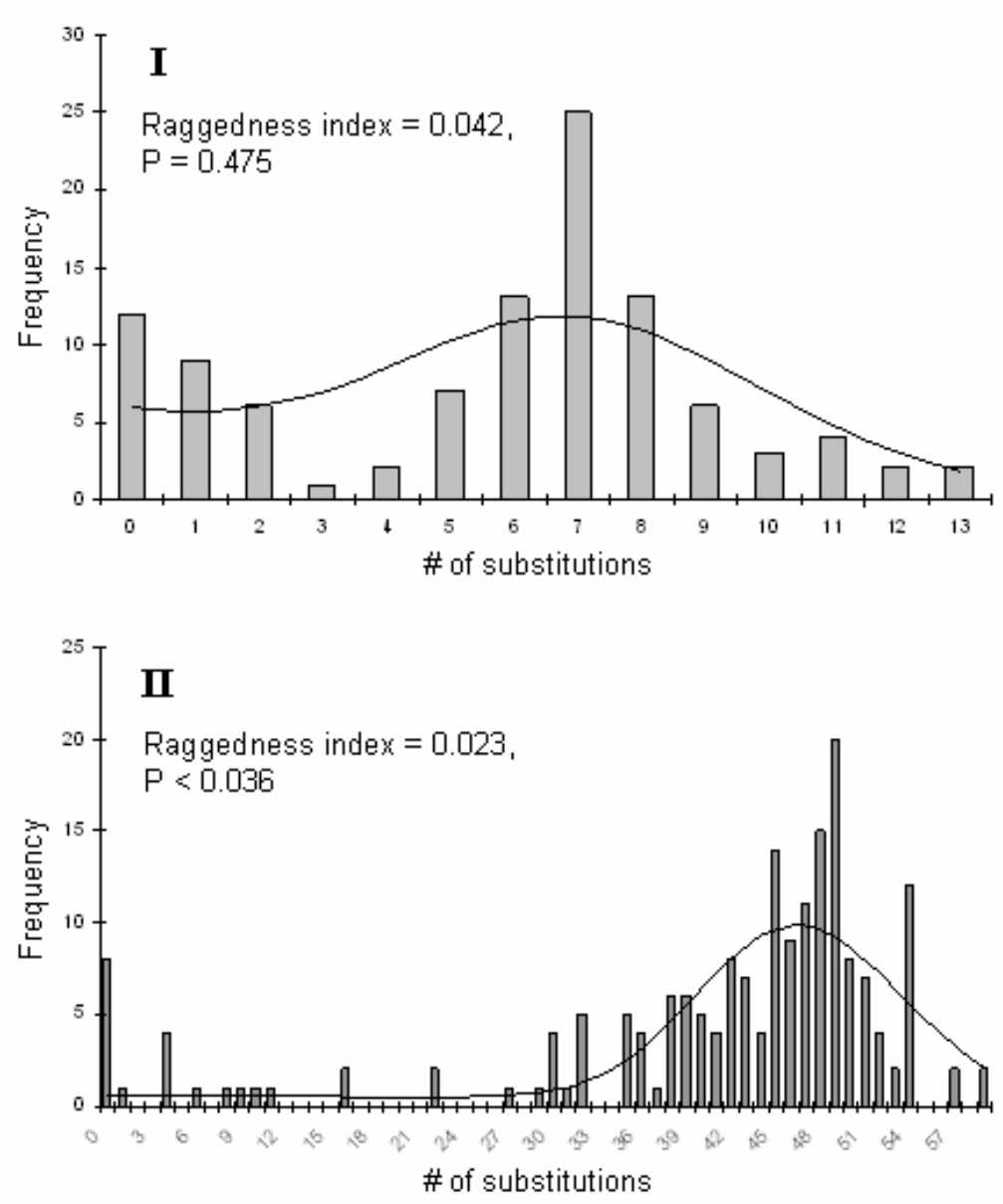


Figure 11 Expected (line) and observed (bars) mismatch distribution given a null distribution compatible with sudden population expansion for (I) haplogroup A and (II) haplogroup B. Harpending's raggedness index (Rogers and Harpending 1992) with significance levels are indicated for each of the two haplogroups.

with a historical population expansion. However, the distribution is clearly unimodal if the three bars representing 0, 1 and 2 differences between haplotypes are excluded. Ruling out these three bars on the left side of the histogram may be justified in light of the results given by the MDIV program, i.e. the extremely low or zero migration rates found in all pairwise species comparisons. In a comparison of haplotypes belonging to three genetically isolated endemic species, high frequencies of no or only a few haplotype differences, reflecting intraspecific variation, are actually what we would expect. Thus, excluding the three bars on the left side in the histogram makes the distribution more or less perfectly unimodal, a pattern which is consistent with a historical demographic expansion. Moreover, I found the raggedness index to be 0.042 (P = 0.475). The sum of squared deviations was < 0.031 (P = 0.228). Given an estimated tau of 7.766 I therefore estimated the population expansion to have happened 684,832 years ago (95% confidence interval: 307,672 – 1,090,829).

In contrast, the mismatch distribution of haplogroup B (*H. horatosphaga, H. meruensis, H. tenera, L. inermis* and *L. new*) is clearly multimodal (Figure 11, II), suggesting that this haplogroup comprises several genetically distinct subgroups. The raggedness index does not contradict this interpretation (raggedness index < 0.023, P < 0.036).

Synonymous versus Non-Synonymous Substitutions

The ratios of non-synonymous versus synonymous substitutions (dn/ds) in the groups analyzed are shown in Table 3. All the ratios are clearly different and especially the group comprising *H. horatosphaga, H. meruensis, H. tenera, L. inermis* and *L. new* (the savannah group in Table 3) has an extremely low ratio compared to the other two groups analyzed. Not surprisingly then, the two-tailed 99% confidence intervals computed in R showed that all ratios were significantly different from each other, as none of them contained the number zero (0.0044 - 0.0173 between *Monticolaria* and *H. nou*, *H. montivaga* and *H. hanangensis*; 0.0143 - 0.0220 between *Monticolaria* and the savannah group; 0.0126 - 0.023 between the savannah group and *H. nou*, *H. montivaga* and *H. hanangensis*).

Table 3 The average number of synonymous substitutions per synonymous site (Pi(s)), the average number of non-synonymous substitutions per non-synonymous site (Pi(a)) and the average ratios of non-synonymous versus synonymous substitutions (dn/ds) for the three groups analyzed. In the table, the savannah group refers to the species *H. horatosphaga*, *H. meruensis*, *H. tenera*, *L. inermis* and *L. new*. All ratios are significantly different from each other (using two-tailed 99% confidence intervals).

	Pi(s)	Pi(a)	dn / ds
Savannah group	0.27982	0.00025	0.00089343
Monticolaria	0.04269	0.00034	0.00796439
H. nou, H. montivaga & *H. hanangensis*	0.19729	0.00467	0.02367074

DISCUSSION

Phylogeography and Speciation

The species within each of the monophyletic groups in the evolutionary trees share roughly the same habitats. *Monticolaria* and the clade comprising *H. nou, H. montivaga* and *H. hanangensis* contain forest dependent species only, while the group comprising *P. clavigera, P. uguenoensis* and *H. regularis* consists solely of savannah inhabitants. In the group consisting of *H. horatosphaga, H. meruensis, H. tenera, L. inermis*, *L. new*, *H. sabuk* and *H. parensis,* nearly all are savannah inhabitants. The only exceptions are *H. sabuk* and *H. parensis* which are forest dependent species.

The forest dependent genus *Monticolaria* is very distinct both morphologically and genetically and appears as the ancestral group in the phylogeny. Thus, savannah inhabiting lineages seem to have evolved from forest living species. The genetic distances between *Monticolaria* and the other lineages confirm an old common ancestor of the ingroup as a whole. Even the common ancestor of the three remaining monophyletic clades seems to be quite ancient, as is indicated by the large genetic distances between these three groups. It is therefore interesting, when considering the old age of the ingroup, that nearly all the speciation events within the four monophyletic clades seem to roughly distribute themselves in two distinct and relatively short periods of time.

Spread of Grasslands and Adaptive Radiation of Savannah Inhabiting Lineages

Some parts of the phylogenetic trees show weak clade support, i.e. bootstrap values and posterior probabilities around 50% and 0.50, respectively. Such low support in molecular phylogenetic analysis is often interpreted as a signal of insufficient data, inappropriate choice of genes or both. However, low clade support might also indicate a radiation of new species over a relative short period of time. The molecular clock approach suggests that several different lineages originated between 6 and 3.5 million years ago. Interestingly, most of these new species appear to have been savannah inhabiting lineages. Neglecting the possibility of extinction for a moment, very few savannah

species seem to have existed seven million years ago. Yet, the number of savannah lineages increased drastically during the late Miocene and early Pliocene.

Based on isotopic signatures of $\delta^{13}C$ from carbonate paleosols and other sources, it appears that there was a substantial increase in C4 grasslands in widely separate parts of the globe from 7 to 4 million years ago (Cerling *et al.* 1993; Morgan *et al.* 1994; Latorre *et al.* 1997; Barry *et al.* 2002). Adaptive speciation to the new niches which this increase of C4 grasslands probably represents therefore seems as an obvious and plausible explanation for the contemporary sudden massive origin of new savannah adopted lineages. However, although primarily driven by vegetation change, other mechanisms were probably also important in the radiation. Further research with larger sample sizes and more populations must be conducted before this radiation can be addressed in greater detail.

Recent Radiations of Forest Dependent Species

All species which belong to the forest dependent clades are trapped in forest refuges at high altitudes on mountains and volcanoes. Further, my analyses revealed that five out of six of these forest dependent species have a quite recent origin. These species also seem to have originated within a very short period of time, as might explain why I am not able to solve the internal relationships between *H. nou, H. montivaga* and *H. hanangensis.* All these results point toward the existence of one or a few recent historical events which played an important role in speciation of the forest dependent species.

From the late Neogene to the late Pleistocene and further on to the present, East Africa has experienced an increased aridification (Trauth *et al.* 2005). Especially during the late Pliocene and Pleistocene African climate became cooler, drier and more seasonal, which triggered the spread of savannah in the eastern parts of the continent (deMenocal 1995). Vegetation shifted from closed canopy to open savannah and evidence suggests a continuous presence of savannah from about 2.5 million years ago and onward (Fernandez and Vrba 2006). However, despite the overall trend of a more arid and cooler climate in this region during the last million years, climatic fluctuations with more humid periods occurred from about 2.7 to 2.5, 1.9 to 1.7 and 1.1 to 0.9 million years ago (deMenocal 1995; Trauth *et al.* 2005). The Plio-Pleistocene climate in the region may

therefore be characterized as a continuum of alternating wet and dry conditions, climatic fluctuations which surely influenced the forest cover in the region. Forests retracted to higher altitudes during dry and cold periods with a consequential increase in the savannah landscape at lower altitudes. This was reversed during warm and wet periods as forests expanded down from the mountains towards lower altitudes, potentially connecting previously isolated forest patches. Accordingly, the periodically cooler and drier East African conditions after about 2.5 million years and their subsequent intensification about 1.7 and 0.9 million years ago may have established opportunities for ecological fragmentation with subsequent genetic isolation of forest dependent species (deMenocal 1995). We might therefore expect more or less simultaneous vicariant speciation events to have happened in forest dependent species as populations were trapped in forest refugia on different mountains in the time periods right after the warm and humid climatic extremes.

The coalescent modeling and the mismatch distribution suggest the three species *H. nou, H. montivaga* and *H. hanangensis*, all of them endemic to respective volcanoes, diverged more or less simultaneously between 500,000 and 700,000 years ago. Additional results from the coalescent analyses indicate that no recent gene flow have happened between the three species. This finding is consistent with a historical isolation scenario. The same coalescent approach suggests that no recent gene flow had happened between *M. meruensis* and *M. hanangensis* either. Further, *M. meruensis* and *M. hanangensis* seem to have diverged about 900,000 years ago, suggesting that these two *Monticolaria* species diverged some hundred thousand years earlier than *H. nou, H. montivaga* and *H. hanangensis* did. However, given the intrinsic uncertainty in all divergence estimates based on molecular markers, it seems reasonable to state that the three *Horatosphaga* species and the two *Monticolaria* species diverged approximately at the same time. Interestingly, all these speciation events seem to have happened right after the last warm and humid maximum in the region about one million years ago. The climatic induced spread of savannah and the consequential creation of forest refuges in the time after this last warm and wet period therefore seems as the primary driving force of speciation in this group of forest dependent species. Yet, even if the two *Monticolaria* species diverged some hundred thousand years earlier than the three *Horatosphaga*

species, the same barrier might still have induced the speciation events in both groups. This is because strict age-equality is not necessarily expected since the time when a forming barrier interrupts gene flow may differ between ecologically different species (Knowlton *et al.* 1993). Although *H. nou, H. montivaga, H. hanangensis* and the two *Monticolaria* species are all forest dependent species, they do not share the same ecological niche. *H. nou, H. montivaga* and *H. hanangensis* are herbaceous vegetation dwellers and need open conditions, while *Monticolaria sp.*, although they also prefer forest edge and clearings, are inhabiting more of the tree and bush layer in forests.

None of the forest dependent species seem to have originated in connection with the two earliest warm and humid maxima about 2.5 and 1.7 million years ago. One possibility is that sufficient reproductive isolating barriers had not evolved between populations when they met during warm and humid periods. Consequentially, populations merged into one species when they mixed during secondary contact. Extinction as a result of competition between ecologically similar (sub)species in warm and wet periods is another possible explanation. However, the lack of species originating right after the two earliest warm and wet maxima may also be a consequence of the low number of species included in this study.

Adaptive speciation to new heterogeneous habitats on geologically new volcanoes represents an alternative explanation to the climate induced vicariant speciation hypothesis. Actually, other studies focusing on evolutionary history in Orthoptera species have found adaptive speciation to play an important role in speciation processes (Trewick and Morgan-Richards 2005). In this study, however, there seems to be a very poor link between the age of the volcanoes and the age of the species which inhabit them. For example, Mt. Kilimanjaro is assumed to have appeared about one million years ago, but *M. kilimandjarica*, endemic to this volcano, seems to be many million years old. Both *M. meruensis* and *M. hanangensis* have more than a 12% genetic distance to *M. kilimandjarica*, suggesting that *Monticolaria* is an old genus. The lineage leading to *M. kilimandjarica* probably originated in the Miocene when the climate in East Africa still favored a continuous forest cover. However, the lineage was probably also affected by the drying climate in Plio-Pleistocene since the species has only been found at Mt. Kilimanjaro. The lineage probably had a more widespread, or at least different,

distribution before the climate caused the forest to retract and finally trapped *M. kilimandjarica* on Mt. Kilimanjaro. Two opposite examples are *H. nou* and *H. hanangensis*, both young species which inhabit forests on geologically old formations. However, adaptive speciation and the vicariant speciation hypothesis are not mutually exclusive. In fact, adaptive speciation may have played an important part in the speciation processes reported here, as both drift and adaptation can cause two independently evolving populations to develop reproductive isolation between each other (Coyne and Orr 2004).

Different Selection Regimes in Savannah versus Forest Habitats?

The presence of continuous forest at lower altitudes in East Africa during the Miocene indicates that forest inhabiting species have existed in the region for a long time. My genetic data also show that the forest dependent lineages are older than species inhabiting the savannah. An obvious suggestion is therefore that the transitions between habitats have primarily been from forest to savannah. To make this transition possible, mutations allowing lineages to utilize new types of food and adapt to novel microclimates must have been acquired. In that respect, a relaxation of selective constraints normally operating on protein coding genes might have been necessary for the transitions to be completed.

One of the mtDNA segments (COI) analyzed in this study is essential in cellular energetics and therefore potentially very important when a taxon acquires novel adaptations to new niches with different dietary conditions (Ballard and Whitlock 2004). If the transition from a forest habitat to a savannah habitat required relaxed selective constrains, I would therefore expect lower rates of non-synonymous to synonymous mutations (dn/ds) in forest dependent species relative to savannah living species in the COI gene. However, the average ratios of dn/ds in the groups analyzed did not support the hypothesis of relaxation of selective constraints in the savannah species. Rather the opposite seems much more likely, i.e. the savannah inhabitants show extremely low dn/ds ratios compared to the forest dependent species clades. This is also confirmed by the confidence intervals computed in R.

The low dn/ds ratio in the savannah group is a result of evolutionary processes occurring over a time span of at least 5 million years. Accordingly, there should have been plenty of time for non-synonymous substitutions to have happened. Indeed, the group comprising the three species *H. nou*, *H. montivaga* and *H. hanangensis* shows a dn/ds ratio about 25 times higher than the savannah group, indicating that extremely strong selective constraints have operated in the savannah lineages. One explanation can be that competition for food and patches with the right microclimate might be more extensive on the savannah compared to the forests. Accordingly, every deviation from the niche which a savannah species is exploiting most effectively will be removed by natural selection. Interestingly in this respect, not allowing non-synonymous changes seems like a consistent trend within all the genetically distinct savannah lineages analyzed. However, these selective constraints have apparently not operated on the morphological level to the same extent, since many of the species belonging to the paraphyletic savannah group have large phenotypic plasticity and pronounced sexual dimorphism (Ragge 1960). These opposite evolutionary patterns concerning external morphology and the evolution of genes important in cell metabolism confirm the mosaic nature of evolutionary change, i.e. organisms (species) do not evolve as a whole, but piecemeal.

Phylogenetic Reconstruction and Taxonomy

The examined species fall into three easily distinguishable groups in addition to the divergent genus *Monticolaria*. These groups do not match the currently used taxonomy at all and none of the three existing genera *Horatosphaga, Peronura* and *Lamecosoma* are monophyletic. Earlier revisions of the Acrometopae species based on morphological characters have pointed out that certain genera, and especially the genus *Horatosphaga*, contain a rather heterogeneous assemblage of species and that new genera should be created if supported by additional data (Ragge 1960). I will therefore argue in favor of a taxonomic revision of some of the groups included in this study, both at the genera and species levels.

The three *Monticolaria* species clustering together was expected since they share several morphological characters which are not found in the other taxa in the ingroup.

Further, *Monticolaria* is not a genus included in Acrometopae. In this respect, this genus also functions as an outgroup to the rest of the taxa included in the study.

The strong support for monophyly of the species *H. montivaga, H. nou* and *H. hanangensis* was not a surprise either as these three species share several ecological and morphological characteristics.. This group has also got relative high genetic distances to the other monophyletic clades within the ingroup. I therefore suggest it would be advisable to create a new genus for *H. hanangensis, H. montivaga* and *H. nou.* Nearly the whole ingroup needs to be revised taxonomically in light of what the phylogenetic analysis have revealed of inaccurate taxonomic classifications. The creation of a new genus for this group of species might therefore be one step in the right direction in order to get a taxonomy that better reflects the evolutionary history of the group.

The close association between *P. uguenoensis*, *P. clavigera* and *H. regularis* in all phylogenetic trees was not anticipated. As is evident from the species names, *H. regularis* is currently believed to belong to a different genus than the two *Peronura* species. Anyhow, my results clearly suggest that *H. regularis* should be placed in the genus *Peronura*, which then will constitute a monophyletic group. Further, the data reveal that the *H. regularis* samples contain two genetically distinct clades. The genetic distance between these two groups corresponds to a period without gene flow lasting for about five million years. This might indicate the existence of two species instead of one. The morphology of these two genetically distinct clades, however, does not indicate the presence of two different species. Although unusual, other studies have also found a genetic differentiation of a similar magnitude within Orthoptera species (Trewick *et al.* 2000; Spooner and Ritchie 2006). Crossing experiments and fitness estimates of the resulting hybrids may reveal whether the clade consists of a pair of sibling species or not.

The last distinguishable group from my phylogenetic trees is the clade containing the savannah species *H. tenera*, *H. heteromorpha*, *H. meruensis*, *L. new* and *L. inermis* (Figure 12). The genus *Lamecosoma* is not monophyletic since samples belonging to this genus cluster together with samples belonging to *Horatosphaga* species. Accordingly, the *Horatosphaga* genus is not monophyletic either. Even at the species level, *H. heteromorpha* and *L. inermis* do not form monophyletic clades. Accordingly, this savannah group contains paraphyly on both species and genera level.

Such inconsistencies of the traditional taxonomy have long been anticipated (Ragge 1960), but can be interpreted in different ways. Large phenotypic plasticity and extensive sexual dimorphism characterize the genus *Horatosphaga* (Ragge 1960). Misclassifying of specimens might therefore be a possibility. A more likely explanation, strongly suggested by the present phylogenetic analyses, the mismatch distribution and the genetic distances computed, is that this savannah group comprises several genetically distinct populations/species which have been separated for varying time spans. However, this conclusion is heavily contradicted by the morphological data which suggest the current taxonomy represents real species boundaries. Anyhow, it is still possible that the currently used taxonomy and the genetic data presented here are compatible. This will be the situation if and only if the lineages causing the paraphyly all descend from the same huge gene pool from which many old lineages have survived to the present. Fortunately,

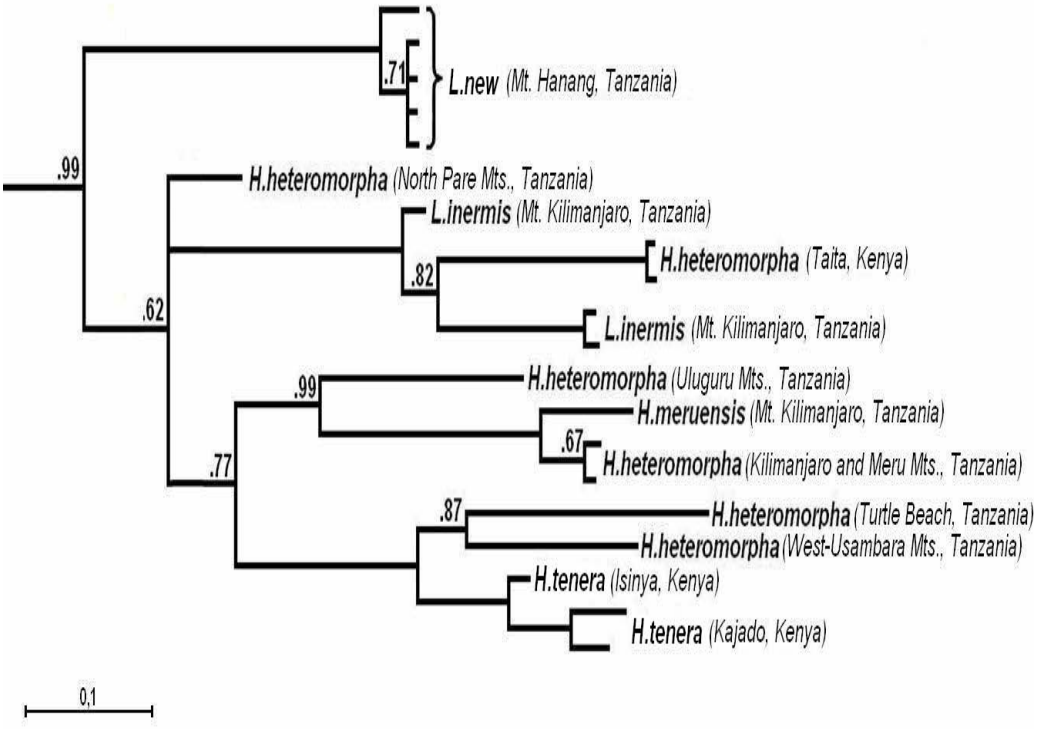


Figure 12 A section of the phylogenetic tree computed using Bayesian inference (BI) analysis based on the COI-gene (see figure 6 for the whole BI tree). Paraphyly on both genus and species level exists in this group of savannah inhabiting lineages. The geographic locations of the different samples are put behind the species names.

this hypothesis will be easy to test. If two different genetic lineages belonging to the same species defined by the current taxonomy are not able to produce viable and fertile offspring, this would falsify both the hypothesis and the current taxonomy. Further studies must therefore be conducted before their taxonomy can be clarified.

CONCLUSIONS

Despite the existence of several convincing examples of single vicariant speciation events, it has been hard to find reliable evidence that vicariant speciation is pervasive in a group of species (Coyne and Orr 2004). To show this, reliable phylogenies for the existence of sister species, accurate molecular clocks and strong evidence for geographic barriers are needed. All these requirements are either met or have the possibility to be met in the group of Orthoptera species analyzed herein. Moreover, younger species are more likely to reflect the biogeography of speciation than older species. Therefore, and despite the quite small sample sizes some of my conclusions build upon, the forest dependent species that originated in a putative radiation about 800,000 years ago seem like good candidates for investigating how climatically induced barriers functioned as a promoter of multiple and simultaneous vicariant speciation events. Consequentially, the continental archipelago in East Africa has proven itself a suitable model system for doing traditional phylogenetic research and studying how fluctuations in climate have affected the evolution of certain segments of the flora and fauna in the region. I suggest that other available forest-living *Horatosphaga* species should be incorporated in future studies. This will allow for additional possibilities to further broaden our understanding of the evolutionary consequences of the fluctuating climate in the region. I also found signals pointing towards an earlier radiation starting around 6 million years ago and lasting about 2 million years. Notably, this radiation overlaps in time with the assumed global spread of C4 grassland, which can be interpreted as adaptive radiation into an emerging savannah habitat. Interestingly, it seems that the savannah inhabiting species are selectively more constrained than their forest-dwelling ancestors, possibly indicating narrower ecological niches on the savannah for bush crickets. Finally, the study clearly demonstrates the need for a taxonomic revision of the ingroup as a whole.

LITERATURE CITED

Akaike, H. (1973). "Information Theory and an Extension of the Maximum Likelihood Principle." Academiai Kiado.

Allegrucci, G., V. Todisco and V. Sbordoni (2005). "Molecular Phylogeography of *Dolichopoda* Cave Crickets (Orthoptera, Rhaphidophoridae): A Scenario Suggested by Mitochondrial DNA." Molecular Phylogenetics and Evolution **37**(1): 153-164.

Avise, J. C. (1998). "The History and Purview of Phylogeography: A Personal Reflection." Molecular Ecology **7**(4): 371-379.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb and N. C. Saunders (1987). "Intraspecific Phylogeography - The Mitochondrial-DNA Bridge between Population-Genetics and Systematics." Annual Review of Ecology and Systematics **18**: 489-522.

Ballard, J. W. O. and M. C. Whitlock (2004). "The Incomplete Natural History of Mitochondria." Molecular Ecology **13**(4): 729-744.

Barry, J. C., M. L. E. Morgan, L. J. Flynn, D. Pilbeam, A. K. Behrensmeyer, S. M. Raza, I. A. Khan, C. Badgley, J. Hicks and J. Kelley (2002). "Faunal and Environmental Change in the Late Miocene Siwaliks of Northern Pakistan." Paleobiology **28**(2): 1-71.

Brower, A. V. Z. (1994). "Rapid Morphological Radiation and Convergence Among Races of the Butterfly *Heliconius erato* Inferred from Patterns of Mitochondrial-DNA Evolution." Proceedings of the National Academy of Sciences of the United States of America **91**(14): 6491-6495.

Burnham, K. P. and D. R. Anderson (1998). "Model Selection and Inference - A Practical Information-Theoretic Approach." Springer-Verlag New York Inc.

Cerling, T. E., Y. Wang and J. Quade (1993). "Expansion of C4 Ecosystems as an Indicator of Global Ecological Change in the Late Miocene." Nature **361**(6410): 344-345.

Coyne, J. A. and H. A. Orr (2004). "Speciation." Sinauer Associates Inc.

deMenocal, P. B. (1995). "Pliopleistocene African Climate." Science **270**(5233): 53-59.

deMenocal, P. B. (2004). "African Climate Change and Faunal Evolution during the Pliocene-Pleistocene." Earth and Planetary Science Letters **220**(1-2): 3-24.

Dennell, R. (2003). "Dispersal and Colonisation, Long and Short Chronologies: How Continuous is the Early Pleistocene Record for Hominids Outside East Africa?" Journal of Human Evolution **45**(6): 421-440.

Emerson, B. C. and G. M. Hewitt (2005). "Phylogeography." Current Biology **15**(10): R367-R371.

Emerson, B. C. and P. Oromi (2005). "Diversification of the Forest Beetle Genus *Tarphius* on the Canary Islands, and the Evolutionary Origins of Island Endemics." Evolution **59**(3): 586-598.

Farris, J. S. (1970). "Methods for Computing Wagner Trees." Systematic Zoology **19**(1): 83-92.

Farris, J. S., M. Kallersjo, A. G. Kluge and C. Bult (1994). "Testing Significance of Incongruence." Cladistics - the International Journal of the Willi Hennig Society **10**(3): 315-319.

Farris, J. S., M. Kallersjo, A. G. Kluge and C. Bult (1995). "Constructing a Significance Test for Incongruence." Systematic Biology **44**(4): 570-572.

Felsenstein, J. (1981). "Evolutionary Trees from DNA-Sequences - a Maximum-Likelihood Approach." Journal of Molecular Evolution **17**(6): 368-376.

Fernandez, M. H. and E. S. Vrba (2006). "Plio-Pleistocene Climatic Change in the Turkana Basin (East Africa): Evidence from Large Mammal Faunas." Journal of Human Evolution **50**(6): 595-626.

Finlayson, C. (2005). "Biogeography and Evolution of the Genus *Homo*." Trends in Ecology & Evolution **20**(8): 457-463.

Fleischer, R. C., C. E. McIntosh and C. L. Tarr (1998). "Evolution on a Volcanic Conveyor Belt: Using Phylogeographic Reconstructions and K-Ar-Based Ages of the Hawaiian Islands to Estimate Molecular Evolutionary Rates." Molecular Ecology **7**(4): 533-545.

Folmer, O., M. Black, W. Hoeh, R. Lutz and R. Vrijenhoek (1994). "DNA Primers for Amplification of Mitochondrial Cytochrome C Oxidase Subunit I from Diverse Metazoan Invertebrates." Molecular Marine Biology and Biotechnology **3**(5): 294-299.

Gillespie, R. (2004). "Community Assembly through Adaptive Radiation in Hawaiian Spiders." Science **303**(5656): 356-359.

Hall, T. A. (1999). "BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT." Nucleic Acids Symposium Series **41**: 95-98.

Harpending, H. C. (1994). "Signature of Ancient Population-Growth in a Low-Resolution Mitochondrial-DNA Mismatch Distribution." Human Biology **66**(4): 591-600.

Hemp, A. (2002a). "Ecology of the Pteridophytes on the Southern Slopes of Mt. Kilimanjaro - I. Altitudinal Distribution." Plant Ecology **159**(2): 211-239.

Hemp, C. (2002b). "New Acrometopae from East Africa (Tettigoniidae: Phaneropterinae)." Journal of Orthoptera Research **11**(1): 67-76.

Hemp, C. and Hemp, A. (2003). "Saltatoria Coenoses of High-Altitude Grassland on Mt. Kilimanjaro, Tanzania (Orthoptera: Saltatoria)." Ecotropica **9**: 71-97.

Hemp, C. (2005). "The Influence of Fire on Saltatoria Diversity in Costal Habitats Near Pangani, Tanzania (East Africa)." Ecotropica **11**: 53-61.

Hemp, C. (2006). "Two New Species of *Horatosphaga* Schaum 1853 from the Highlands of East Africa (Tettigoniidae: Phaneropterinae)." Journal of Orthoptera Research **15**(2).

Huelsenbeck, J. P. and F. Ronquist (2001). "MRBAYES: Bayesian Inference of Phylogenetic Trees." Bioinformatics **17**(8): 754-755.

Juan, C., P. Oromi and G. M. Hewitt (1995). "Mitochondrial-DNA Phylogeny and Sequential Colonization of Canary-Islands by Darkling Beetles of the Genus *Pimelia* (Tenebrionidae)." Proceedings of the Royal Society of London Series B-Biological Sciences **261**(1361): 173-180.

Juan, C., P. Oromi and G. M. Hewitt (1996). "Phylogeny of the Genus *Hegeter* (Tenebrionidae, Coleoptera) and its Colonization of the Canary Islands Deduced from Cytochrome Oxidase I Mitochondrial DNA Sequences." Heredity **76**: 392-403.

Knowlton, N., L. A. Weigt, L. A. Solorzano, D. K. Mills and E. Bermingham (1993). "Divergence in Proteins, Mitochondrial-DNA, and Reproductive Compatibility across the Isthmus of Panama." Science **260**(5114): 1629-1632.

Kumar, S., K. Tamura and M. Nei (2004). "MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment." Briefings in Bioinformatics **5**(2): 150-163.

Latorre, C., J. Quade and W. C. McIntosh (1997). "The Expansion of C-4 Grasses and Global Change in the Late Miocene: Stable Isotope Evidence from the Americas." Earth and Planetary Science Letters **146**(1-2): 83-96.

Lovett, J. C. and S. K. Wasser (1993). "Biogeography and Ecology of the Rain Forests of Eastern Africa." Cambridge University Press.

Morgan, M. E., J. D. Kingston and B. D. Marino (1994). "Carbon Isotopic Evidence for the Emergence of C4 Plants in the Neogene from Pakistan and Kenya." Nature **367**(6459): 162-165.

Nielsen, R. and J. Wakeley (2001). "Distinguishing Migration from Isolation: A Markov Chain Monte Carlo Approach." Genetics **158**(2): 885-896.

Pestano, J., R. P. Brown, N. M. Suarez, J. Benzal and S. Fajardo (2003). "Intraspecific Evolution of Canary Island Plecotine Bats, Based on mtDNA Sequences." Heredity **90**(4): 302-307.

Posada, D. and K. A. Crandall (1998). "MODELTEST: Testing the Model of DNA Substitution." Bioinformatics **14**(9): 817-818.

Ragge, D. R. (1960). "The Acrometopae of the Ethiopean Region: A Revision, with Notes on the Sexual Dimorphism Shown by the Group (Ortoptera: Tettigoniidae)." Bulletin of the British Museum (Natural History) **8**(7): 267-333.

R Development Core Team (2006). "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria.

Rogers, A. R. and H. Harpending (1992). "Population-Growth Makes Waves in the Distribution of Pairwise Genetic-Differences." Molecular Biology and Evolution **9**(3): 552-569.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer and R. Rozas (2003). "DnaSP, DNA Polymorphism Analyses by the Coalescent and Other Methods." Bioinformatics **19**(18): 2496-2497.

Saitou, N. and M. Nei (1987). "The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees." Molecular Biology and Evolution **4**(4): 406-425.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). "Molecular Cloning: A Laboratory Manual." New York, Cold Spring Harbour Lab. Press

Schneider, S. and L. Excoffier (1999). "Estimation of Past Demographic Parameters from the Distribution of Pairwise Differences when the Mutation Rates Vary among Sites: Application to Human Mitochondrial DNA." Genetics **152**(3): 1079-1089.

Schneider, S., D. Roessli and L. Excoffier (2000). "Arlequin ver. 2.000: A Software for Population Genetics Data Analysis." Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu and P. Flook (1994). "Evolution, Weighting, and Phylogenetic Utility of Mitochondrial Gene-Sequences and a Compilation of Conserved Polymerase Chain-Reaction Primers." Annals of the Entomological Society of America **87**(6): 651-701.

Slatkin, M. and R. R. Hudson (1991). "Pairwise Comparisons of Mitochondrial-DNA Sequences in Stable and Exponentially Growing Populations." Genetics **129**(2): 555-562.

Spooner, L. J. and M. G. Ritchie (2006). "An Unusual Phylogeography in the Bushcricket *Ephippiger ephippiger* from Southern France." Heredity **97**(6): 398-408.

Sturmbauer, C., S. Koblmuller, K. M. Sefc and N. Duftner (2005). "Phylogeographic History of the Genus *Tropheus*, a Lineage of Rock-Dwelling Cichlid Fishes Endemic to Lake Tanganyika." Hydrobiologia **542**: 335-366.

Swofford, D. L. (2001). "PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)." Sinauer Associates Inc.

Tajima, F. (1993). "Simple Methods for Testing the Molecular Evolutionary Clock Hypothesis." Genetics **135**(2): 599-607.

Tonnis, B., P. R. Grant, B. R. Grant and K. Petren (2005). "Habitat Selection and Ecological Speciation in Galapagos Warbler Finches (*Certhidea olivacea* and *Certhidea fusca*)." Proceedings of the Royal Society of London Series B-Biological Sciences **272**(1565): 819-826.

Trauth, M. H., M. A. Maslin, A. Deino and M. R. Strecker (2005). "Late Cenozoic Moisture History of East Africa." Science **309**(5743): 2051-2053.

Trewick, S. A. (2000). "Molecular Evidence for Dispersal Rather than Vicariance as the Origin of Flightless Insect Species on the Chatham Islands, New Zealand." Journal of Biogeography **27**(5): 1189-1200.

Trewick, S. A. and M. Morgan-Richards (2005). "After the Deluge: Mitochondrial DNA Indicates Miocene Radiation and Pliocene Adaptation of Tree and Giant Weta (Orthoptera : Anostostomatidae)." Journal of Biogeography **32**(2): 295-309.

Trewick, S. A., G. P. Wallis and M. Morgan-Richards (2000). "Phylogeographical Pattern Correlates with Pliocene Mountain Building in the Alpine Scree Weta (Orthoptera, Anostostomatidae)." Molecular Ecology **9**(6): 657-666.

Wetzer, R. (2002). "Mitochondrial Genes and Isopod Phylogeny (Peracarida: Isopoda)." Journal of Crustacean Biology **22**(1): 1-14.

APPENDICES

Appendix 1 - Coalescent Analyses

In Nielsen and Wakeley's (2001) method, a coalescent model which employs Bayesian inference in addition to a likelihood framework is established using a Markov chain Monte Carlo (MCMC) method to search parameter space while jointly estimating three parameters: Theta (θ) ($2N_e\mu$, i.e. two times the effective population size times the mutation rate), T (the divergence time between the two populations scaled by the population size) and M (the migration rate between the two populations, also scaled by the population size). Inference regarding the parameters is based on the posterior probability distribution of the parameters, $p(T, M, \theta \mid x)$, where x is the DNA sequence data using uniform priors. However, the parameter space needs to be constrained to ensure that the posterior distributions are appropriate, i.e. biological plausible. The maximum values for the parameters to be estimated must therefore be defined prior to the analysis. Parameter estimates close to the prior maximum value chosen strongly indicate a preset maximum value too small. Accordingly, I ran the analysis again with a higher prior max value for the parameter in question if this was found. I also conducted reruns of some of the analysis with different preset maximum values to check if alternative maximum values affected the posterior distribution of parameter

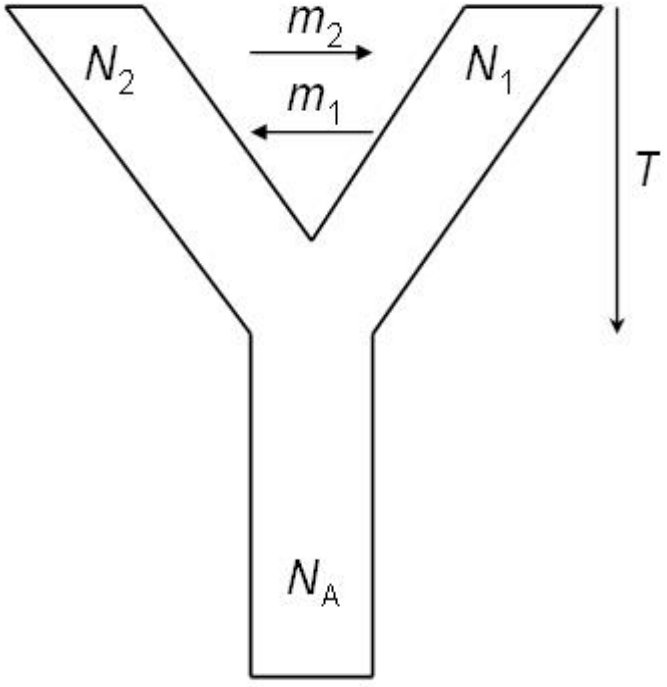


Figure A1-1 A graphical representation of the model used in MDIV**.**

estimates. None of these reruns indicated that the a priori maximum value had any effect on the estimated parameter value. The model in MDIV assumes that two populations arose from a single ancestral population at generation t in the past. Further, migration between the two populations are set to be equal ($m_1 = m_2 = M$) and all three population sizes are equal and constant ($N_1 = N_2 = N_A$) (Figure A1-1).

I conducted four pairwise comparisons using COI sequences on five different species. Species comparisons were only conducted within monophyletic groups, and only on species pairs that were separated by relatively low genetic distances. θ, M and T were estimated using 5,000,000 generations of MCMC with 500,000 burn in cycles. For each pairwise comparison, a minimum of three chains with different random seeds were run using the following preset maximum values: Mmax = 1 or 10 and Tmax = 10, 20 or 30. The default theta-maximum values suggested by MDIV were always preferred. θ, M and T were all estimated from the mode of their respective posterior distribution. Confidence intervals were calculated using Akaike's information criteria (AIC: Burnham and Anderson 1998)

In order to use the parameters estimated by MDIV to calculate divergence times, I converted coalescent time units to years as follows:

$$Tdiv = (T \times \theta) \div (2 \times \mu)$$

where μ is equal to the expected number of mutations that will occur in a sample of *n* base pairs per generation.

MDIV also gives a point estimate of the expected time to the most recent common ancestor (TMRCA) of all the sequences given the data. This point estimate is converted to years since the most recent common ancestor as follows:

$$tRMCA = (TRMCA \times \theta) \div (2 \times \mu)$$

Appendix 2 - Mismatch Distributions

A mismatch distribution is the distribution of the observed number of differences between pairs of haplotypes. The shape of the distribution is normally influenced by prehistoric population demography. For example, samples drawn from a population of constant size and at demographic equilibrium normally produce a ragged multimodal distribution of pairwise differences. A different shape is expected in a population that recently passed through a demographic expansion. Population growth causes the retention of sequences that otherwise would have been lost and consequentially very often creates a smooth and unimodal distribution of pairwise differences. DNA sequence diversity may therefore provide an instrument for examining prehistoric demography (Harpending 1994), but can as mentioned in the main text also be used to detect and estimate the timing of recent species radiation.

The mismatch distributions were calculated as implemented in ARLEQUIN v.2.000 (Schneider *et al.* 2000). This program estimates the parameters of the demographic expansion by a generalized nonlinear least-square approach, as proposed by Schneider and Excoffier (1999). The model assumes that a stationary population at equilibrium experiences a sudden expansion in population size from N_0 to N_1 t generations ago after which the population rapidly reaches a new stationary phase. Based on the mean and variance of all the pairwise haplotype differences, the general nonlinear least-square approach is used to estimate the three demographic parameters $\tau = 2\mu t$, $\theta_0 = 2\mu N_0$ and $\theta_1 = 2\mu N_1$, where μ is the mutation rate for the whole haplotype. Approximate confidence intervals for the three parameters are obtained by a parametric bootstrap approach (Schneider and Excoffier 1999). The validity of the estimated step-wise expansion model is tested using the same parametric bootstrap approach, where the sum of square deviation (SSD) between the observed and the expected mismatch distribution is used as a test statistic. Under the hypothesis that the estimated parameters are the true ones, the P-value of the test is calculated as:

$$P = \frac{\textit{number of SSD simulated larger or equal to SSD observed}}{X}$$

where X equals the total number of SSDs simulated. As the equation shows, the validity of the model is confirmed by a non-significant P-value. With a trusted model, the timing of the putative expansion event can be estimated from τ, which represents the expected number of differences between two randomly drawn haplotypes at time t since the population expansion. τ is defined as 2μt and simple rearrangements give

t = (τ differences per haplotype)
/ [(divergence rate per million years * number of nucleotides)
/ 1 million years]

95 % confidence intervals were estimated from the same parametric bootstrap approach that was used to test the validity of the expansion model, using 10.000 replicates.

ARLEQUIN also computes the raggedness index of the observed distribution, as defined by Harpending (1994). This index indicates the shape of the mismatch distribution as it takes larger values for multimodal distributions than for unimodal distributions. A small raggedness index value indicates low raggedness, typical for recently non-stationary expanding populations. The significance of the raggedness value is tested as described above for SSD using 10,000 bootstrap replicates, i.e. a non-significant P-value increases our confidence in the raggedness index value computed.

Appendix 3 - Maximum Parsimony, Neigbor Joining and Maximum Likelihood Phylogenetic Trees

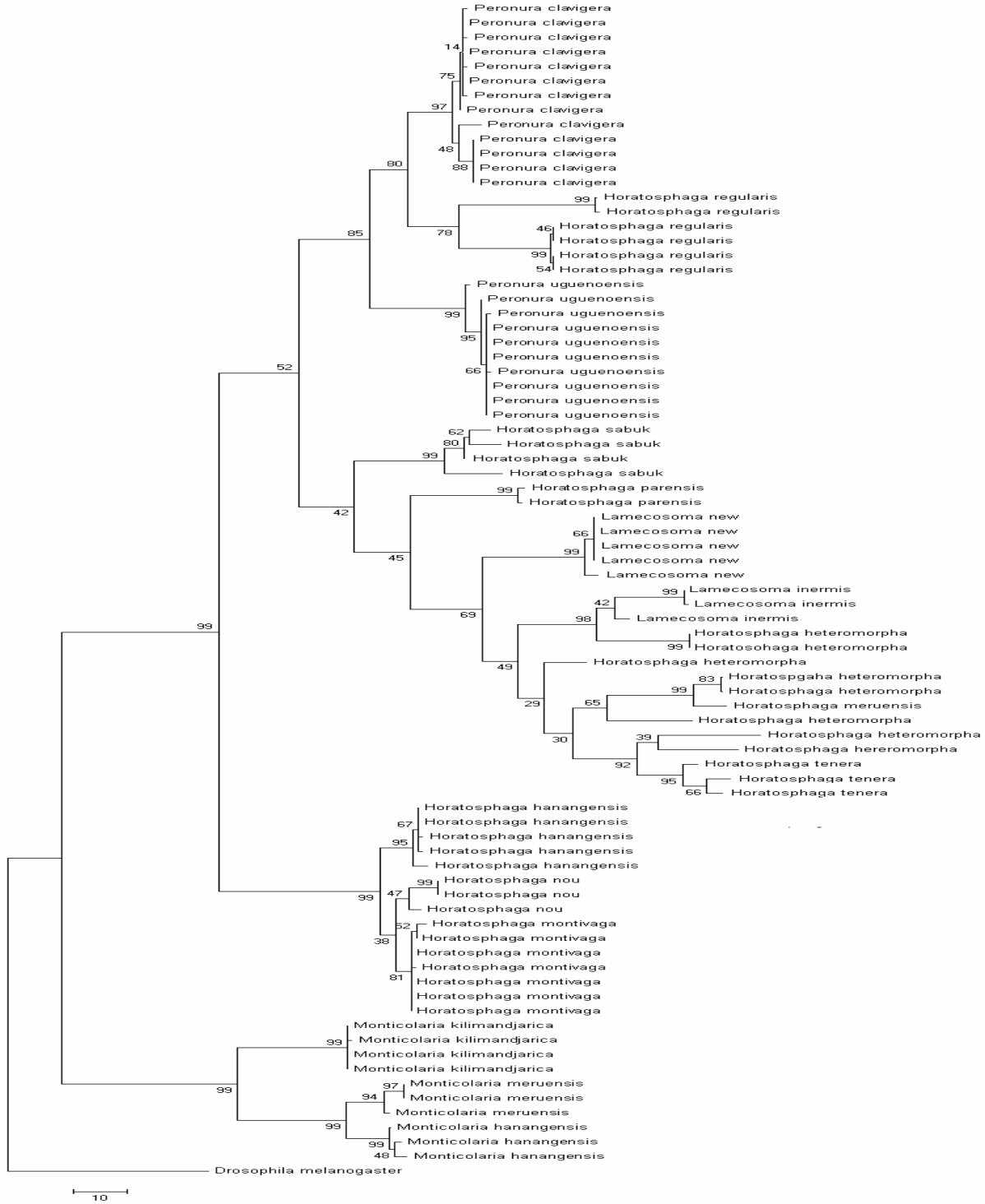


Figure A3-1 COI maximum parsimony tree (bootstrap consensus tree), calculated as described in the "Materials and Methods" section.

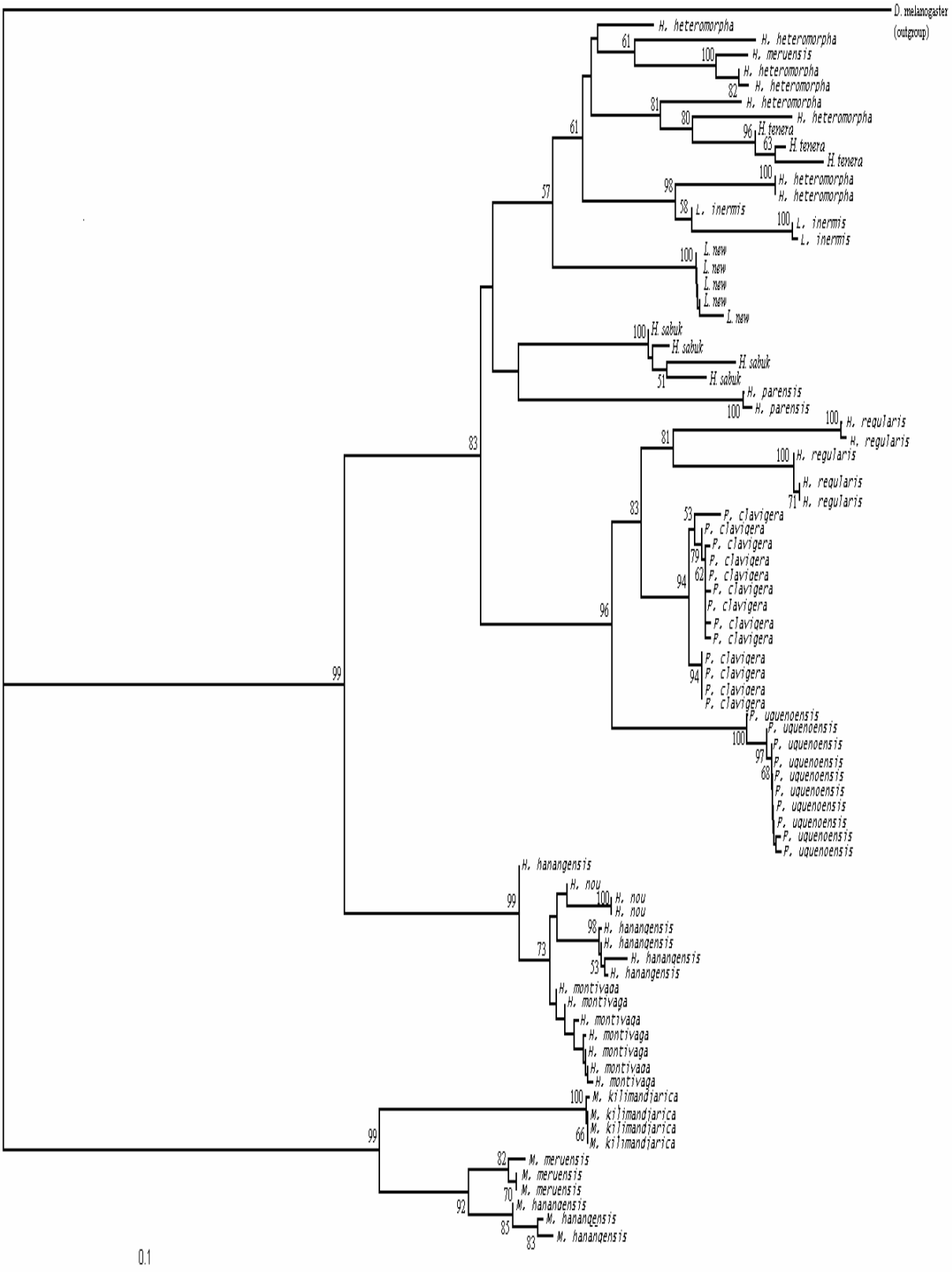


Figure A3-2 COI neighbor joining tree (bootstrap 50% majority-rule consensus tree), calculated as described in the "Materials and Methods" section.

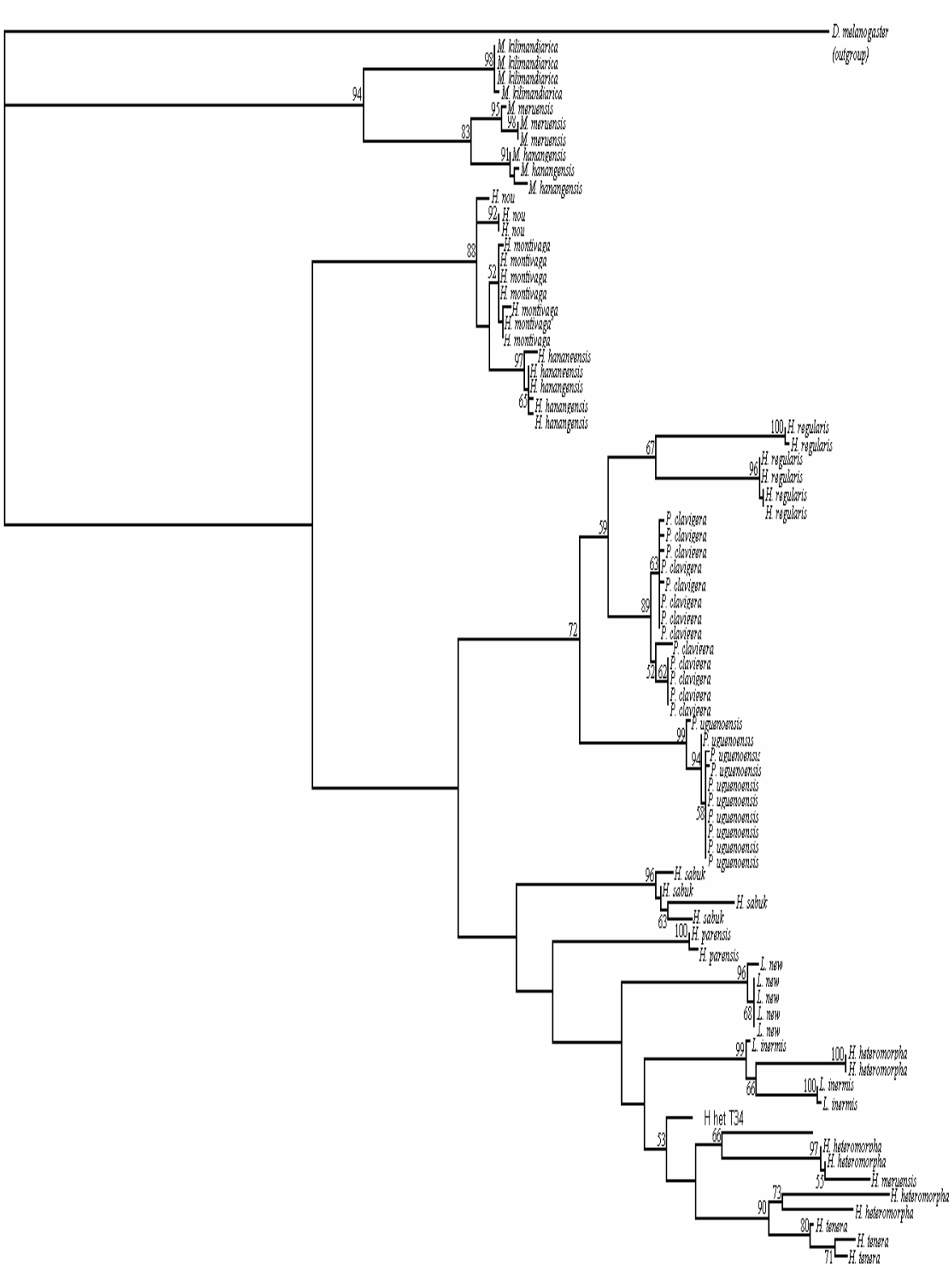


Figure A3-3 COI maximum likelihood tree (bootstrap 50% majority-rule consensus tree), calculated as described in the "Materials and Methods" section.

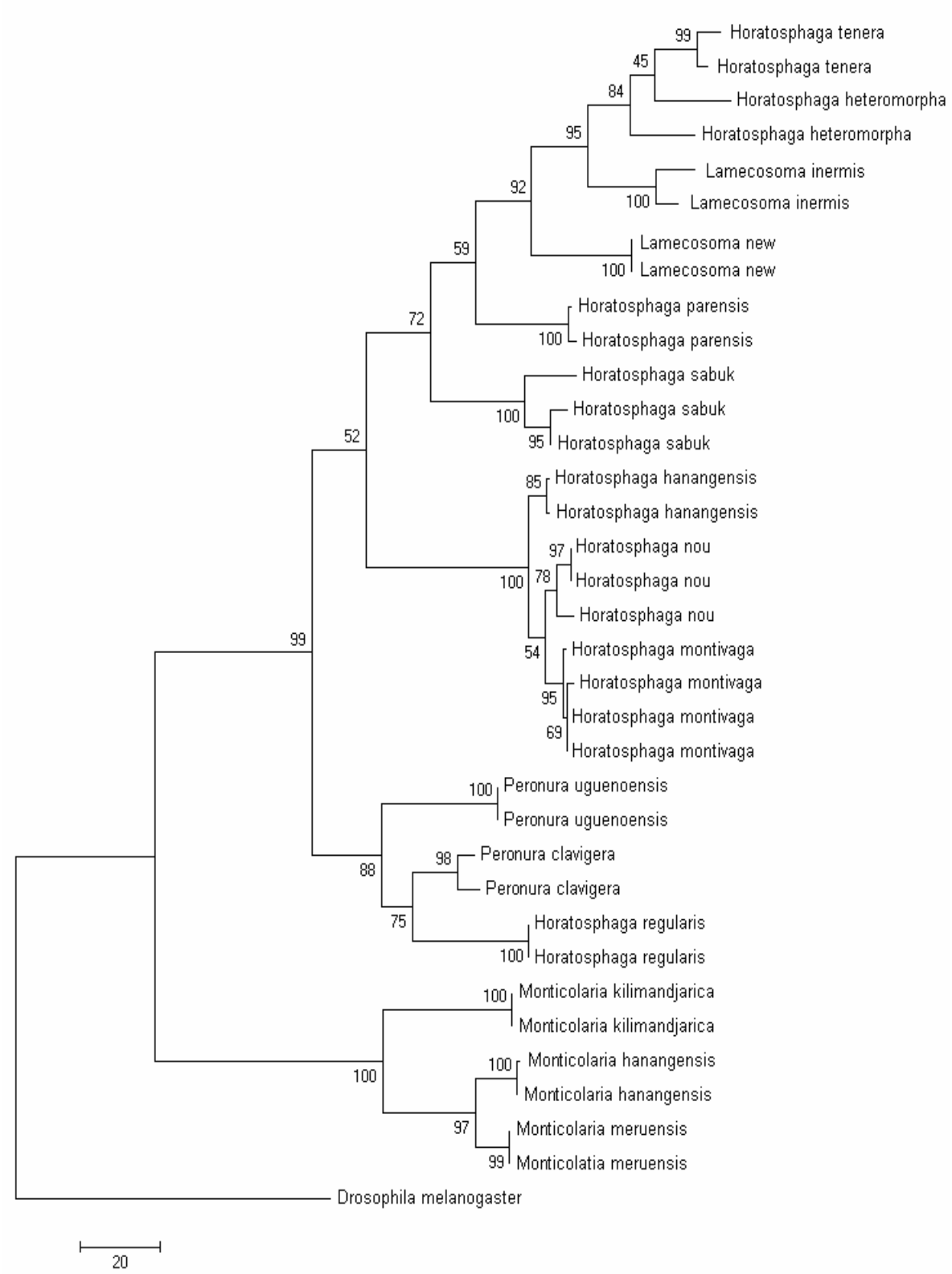


Figure A3-4 COI/12S maximum parsimony tree (bootstrap consensus tree), calculated as described in the "Materials and Methods" section

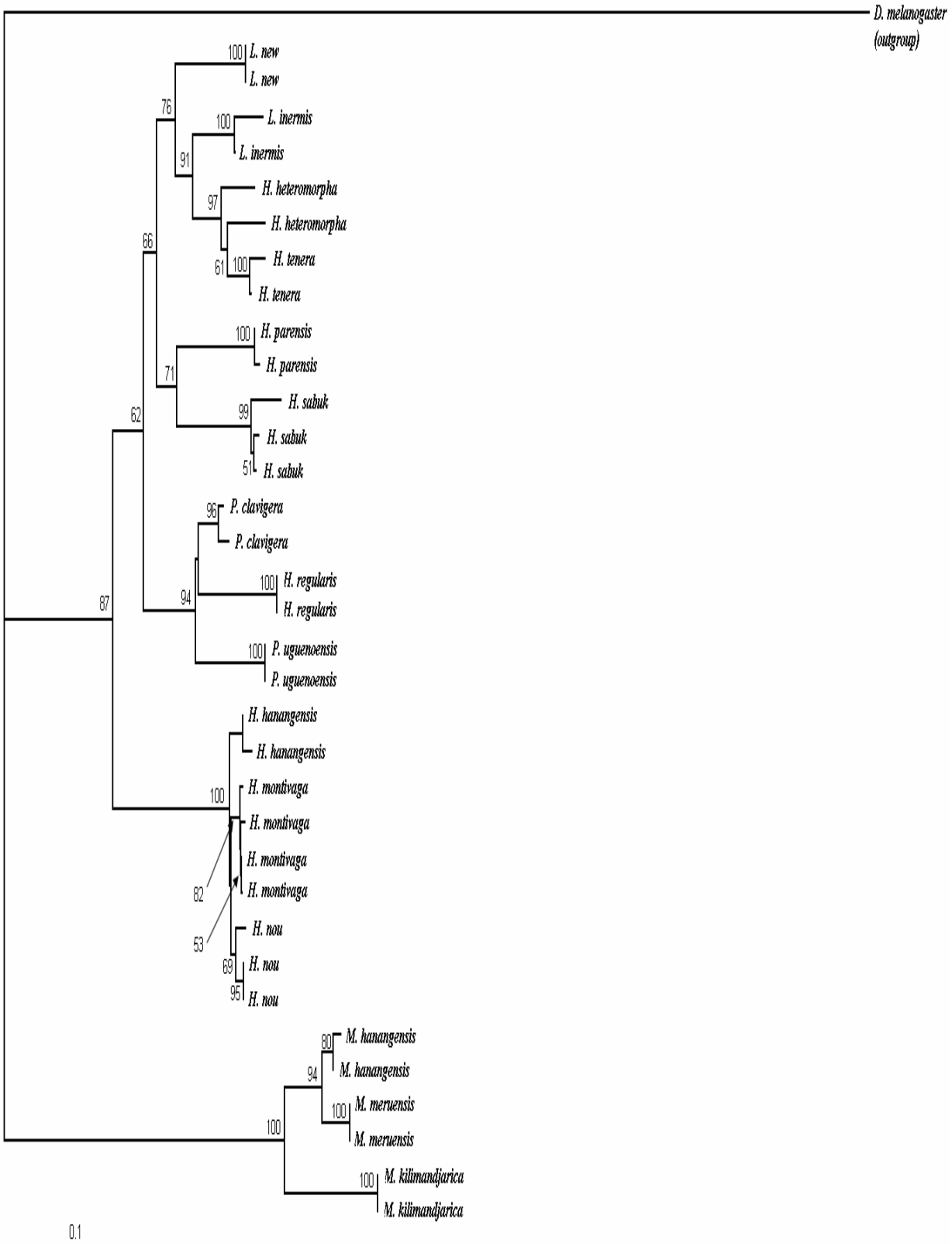


Figure A3-5 COI/12S neighbor joining tree (bootstrap 50% majority-rule consensus tree), calculated as described in the "Materials and Methods" section

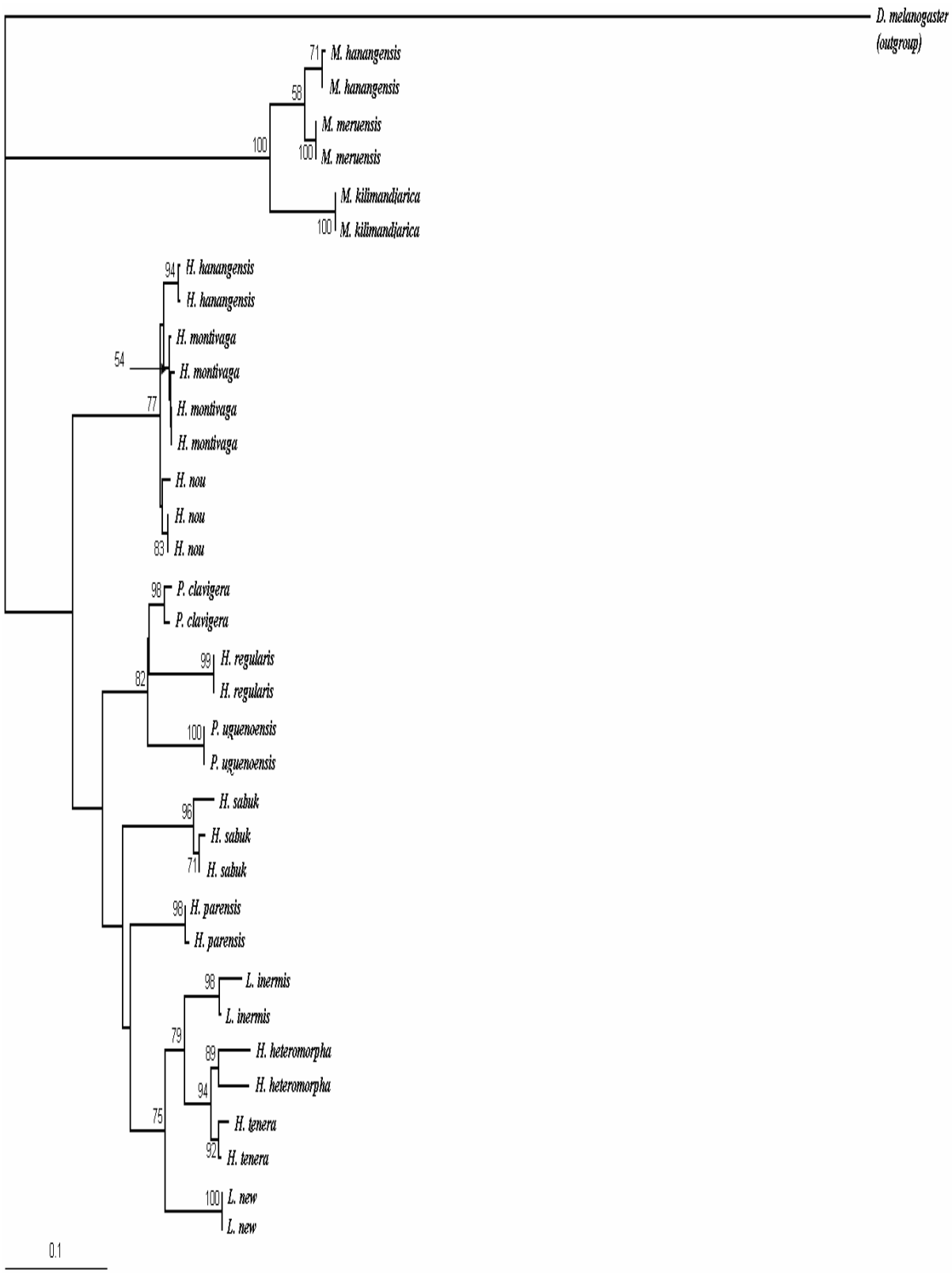


Figure A3-6 COI/12S maximum likelihood tree (bootstrap 50% majority-rule consensus tree), calculated as described in the "Materials and Methods" section

Appendix 4 Pairwise GTR + I + Γ (below diagonal) and uncorrected (p-distances) (above diagonal) genetic distances among consensus cytochrome oxidase subunit I (COI) mitochondrial sequences from all species included in the study.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1)	*L. inermis*	-	0.042	0.106	0.102	0.098	0.093	0.085	0.072	0.110	0.062	0.106	0.095	0.146	0.155	0.151	0.066
(2)	*H. heteromorpha*	0.057	-	0.091	0.087	0.087	0.074	0.068	0.032	0.102	0.042	0.093	0.085	0.153	0.153	0.144	0.043
(3)	*H. nou*	0.253	0.187	-	0.004	0.015	0.102	0.102	0.098	0.117	0.108	0.117	0.102	0.155	0.155	0.149	0.095
(4)	*H. montivaga*	0.226	0.168	0.016	-	0.015	0.098	0.098	0.095	0.117	0.108	0.119	0.102	0.157	0.157	0.151	0.091
(5)	*H. hanangensis*	0.228	0.174	0.029	0.020	-	0.098	0.102	0.095	0.113	0.108	0.117	0.098	0.161	0.159	0.153	0.095
(6)	*H. sabuk*	0.186	0.105	0.217	0.205	0.222	-	0.072	0.091	0.096	0.093	0.098	0.074	0.155	0.147	0.140	0.070
(7)	*H. parensis*	0.163	0.095	0.237	0.211	0.239	0.119	-	0.081	0.112	0.078	0.108	0.093	0.146	0.163	0.149	0.072
(8)	*H. meruensis*	0.120	0.042	0.237	0.207	0.225	0.167	0.153	-	0.104	0.066	0.098	0.093	0.159	0.157	0.147	0.066
(9)	*H. regularis*	0.239	0.187	0.238	0.217	0.223	0.187	0.234	0.210	-	0.110	0.076	0.045	0.144	0.144	0.146	0.102
(10)	*H. tenera*	0.113	0.052	0.237	0.236	0.239	0.157	0.149	0.114	0.232	-	0.104	0.100	0.157	0.164	0.155	0.062
(11)	*P. uguenoensis*	0.211	0.141	0.250	0.251	0.258	0.155	0.200	0.188	0.134	0.194	-	0.059	0.151	0.147	0.151	0.091
(12)	*P. clavigera*	0.191	0.132	0.201	0.183	0.186	0.130	0.160	0.182	0.074	0.198	0.086	-	0.153	0.147	0.147	0.079
(13)	*M. kilmandjarica*	0.436	0.371	0.428	0.436	0.465	0.453	0.451	0.401	0.426	0.390	0.450	0.484	-	0.074	0.074	0.151
(14)	*M. meruensis*	0.440	0.384	0.420	0.397	0.437	0.419	0.438	0.395	0.384	0.433	0.439	0.405	0.121	-	0.026	0.153
(15)	*M. hanangensis*	0.442	0.351	0.418	0.418	0.425	0.393	0.408	0.383	0.418	0.402	0.437	0.415	0.122	0.038	-	0.147
(16)	*L. new*	0.123	0.069	0.221	0.188	0.221	0.141	0.148	0.119	0.214	0.115	0.173	0.155	0.394	0.401	0.382	-